National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

UNIVERSITY OF ALBERTA

A PARAMETER-CONTROLLED RELAXATION ALGORITHM FOR BAYESIAN RESTORATION OF IMAGES USING THE MAXIMUM ENTROPY METHOD (DESIGN, ANALYSIS, MACH BAND SIMULATION AND PARALLELIZATION)

by

KALPAGAM KRISHNAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL ENGINEERING

EDMONTON, ALBERTA

FALL 1990

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-64776-0

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: *KALPAGAM KRISHNAN*

TITLE OF THESIS: *A PARAMETER-CONTROLLED RELAXATION ALGORITHM FOR BAYESIAN RESTORATION OF IMAGES USING THE MAXIMUM ENTROPY METHOD (DESIGN, ANALYSIS, MACHBAND SIMULATION AND PARALLELIZATION)*

DEGREE FOR WHICH THESIS WAS PRESENTED: *Ph.D.,*

YEAR THIS DEGREE GRANTED: *FALL 1990*

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

33, S. N. Das Lay out, street no. 4,
Tatabad, Coimbatore, India 641 012

Date: 5th Oct 1990

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *A PARAMETER-CONTROLLED RELAXATION ALGORITHM FOR BAYESIAN RESTORATION OF IMAGES USING THE MAXIMUM ENTROPY METHOD (DESIGN, ANALYSIS, MACHBAND SIMULATION AND PARALLELIZATION)* submitted by *KALPAGAM KRISHNAN* in partial fulfillment of the requirements for the degree of *DOCTOR OF PHILOSOPHY*.

W. B. Joerg

R. Elliott

O. Kuebler

X. Li

R. E. Rink

Date: 3 Oct 1990

In the inspiring memory of my father Dr. K. R. Krishnan,
to my mother Mrs. Saraswathy Krishnan and
to my husband and best friend, Param.

ABSTRACT

A relaxation scheme is developed for the maximum entropy method and applied to the Bayesian restoration of images. The scheme involves optimizing a system by controlling its states in sequence, using specific parameter strategies. The results evolve directly from the functional analysis of a stationary point equation derived from the Bayesian-based entropy optimization functional. The convergence behaviour of the system is characterized in terms of its state-entropy. Given a degraded input image and the degradation parameter(s), the relaxation scheme performs restoration using a numerical algorithm. For inputs degraded by Gaussian noise, it is shown that the specification of noise variance is not necessary. Using a set of criteria, the algorithm estimates the control parameters adaptively to minimize the influence of the strength of the external constraints upon the system. The algorithm also allows the user to specify a parameter to control the speed of convergence. Test studies with worst case examples demonstrate an expected behaviour of the algorithm along with the performance figures showing an improvement between 58% and 87% over the constrained least squares approach. In a specific test study, the *ME* relaxation algorithm is oberved to simulate the psycho-physical characteristics of Mach bands in biological visual systems. Analytical studies reveal the underlying mechanism similar to Mach's non-linear biological visual model but differs by its response-dependent basis. Test studies show new prospects for the *ME* method in edge detection and enhancement applications. Motivated by the results, parallelization of the restoration algorithm is attempted using two concepts of parallelism: instruction and image domain partitioning. The domain patitioning parallelism is approached with the aim of realizing a VLSI implementation based on dedicated parallel architectures. Initial implementation studies have been conducted using Myrias parallel computers, which are general purpose, MIMD (mutiple instruction and multiple data stream) computers. The performance studies show optimum efficiencies of 91% with 16 processors for convolution algorithm and 78.4% with 8 processors for the maximum entropy deconvolution algorithm using the relaxation scheme.

PREFACE

The term 'entropy', coined from a Greek word meaning 'transformation', has its origins in the field of physical thermodynamics. Today, many areas of science and engineering, ranging from optics and digital communication to modern expert and artificial intelligence systems refer to the term. The intriguing aspect of this growth is the common link through *plogp* and *logp* functional forms (*p* stands for a probability distribution). The origin of this link resides in C. E. Shannon's celebrated work in digital communication. His work demonstrated the unique and special properties of *plogp* form (named as 'entropy' by Shannon) leading to its formulation as a measure of uncertainty in probability distributions. The significance of the uniqueness properties of the form was soon realized in many areas of science and engineering, which led to the wide spread use of entropy concepts. This thesis intends to explore and contribute to the understanding and the usefulness of the entropy concepts in image processing.

The entropy concepts in general show a long history of developments in the scientific domain which is outlined in chapter 1. The maximum entropy (*ME*) method, as the name implies, makes use of entropy maximization concepts for certain types of scientific problems. The method, with growing number of applications and conceptual developments, is subject to many different interpretations and criticisms. As a consequence, the foundation of the method tends to remain obscured and poorly understood. In chapter 2, the method as applied to images is investigated for its foundations from the view points of information theoretic, Bayesian and physical and statistical approaches. The practical advancements of the method are also investigated to make an assessement on the present status of the method.

One of the major obstacles facing the *ME* method in *image processing applications* is understood to be due to its computational complexity. The complexity arises from two factors. Computationally, the method involves a non-linear constrained optimization

problem to obtain a solution. Obtaining an optimal solution is a standard difficulty of any optimization problem. A general convincing algorithm with proven convergence characteristics has not yet emerged for the method. The second complexity factor arises from the high dimensionality of the images and the tediousness of the computations themselves. In technical terms, the algorithmic time complexities are as high as $O(n^4)$; with n typically ranging from 256 x 256 to 1024 x 1024.

Starting from first principles, a procedure is derived, analysed for its merits and computational difficulties, and developed into a practical form. In this approach, the *ME* solution is found to be the result of controlling the states of a system characterized by their entropies. The convergence mechanism of the scheme are evaluated from the analysis results of the scheme. Chapter 3 gives full details of the different stages of design and development. The other complexity factor arising from high dimensionality is a fundamental problem of implementation. The use of parallel processing schemes become indispensable to meet the massive and high-speed computing requirements over large problem sizes. Besides, the study constitutes an exploration phase for the design of dedicated VLSI implementation architectures. Chapter 7 explores the possibilities of parallelizing the developed algorithm using instruction parallelism and domain partitioning concepts. Performance results obtained using Myrias parallel computers are reported.

The numerical and the functional behaviour characteristics of the relaxation scheme are studied in chapter 4 using practical image examples. Comparative evaluations are made wherever possible. Significant observations and inferences had become possible from these test performances. Under specific test conditions, the *ME* estimation results showed marked similarities with the psychophysical characteristics of the biological visual systems. With carefully planned tests and analytical studies, evidences are collected and the results are interpreted in chapter 5. As a part of this study, Mach's biological model of visual enhancement has been analysed in depth in chapter 6. The results show the possibility of using the *ME* method for edge detection and enhancement applications.

The text contains many new terms, particularly in chapter 3, which are denoted by bold styled characters. The terms under emphasis appear in italic font. The text also contains terms within quotes, introduced for effectiveness. When referring to the methods based on the concepts of entropy in general, the term 'methods' is used; its singular form is used when the discussion centers around the specific maximization concept of entropy. As far as possible, distinction is emphasized between the terms method, technique, procedure, and an algorithm. Their respective meanings, as followed in this thesis are outlined in Appendix-1.

The test studies reported in this thesis work were carried out under different computing environments. The associated difficulties in terms of the migration of data and programs have restricted the experimental studies to selected test samples. The choice of test samples (images) have another restriction in terms of the available input-output resources. Appendix-2 shows the various computing systems, input-output resources and the migration paths followed in this thesis work.

ACKNOWLEDGMENTS

I gratefully acknowledge the Faculty of Graduate Studies and Research and the Department of Electrical Engineering for providing me the valuable opportunity and experience of doing a Ph. D program at the University of Alberta. I wish to thank my supervisor, Professor. W. B. Joerg with deep gratitude for his support, encouragement and guidance in carrying out this thesis work. I am particularly thankful for all his efforts in making me understand and work with the stimulating concepts of parallel processing. I also thank Dr. Z. J. Koles of the Department of Applied Sciences in Medicine for his initial guidance. Heartfelt thanks to Dr. P. R. Smy, Dr. A. M. Robinson and Dr. R. P. W. Lawson, whose timely support and advice mean a lot to this thesis work, much beyond words. I wish to thank Dr. R. E. Rink, Dr. R. J. Elliott, Dr. X. Li and Dr. C. G. Englefield for their valuable suggestions, discussions and encouragement, which have contributed immensely to the progress and satisfactory completion of this thesis work. I sincerely thank my external examiner, Dr. Kuebler for his valuable comments and his time in reviewing my thesis. I also wish to make a special mention of the seminars of his group at ETH, which provided a valuable learning experience for this thesis work.

The services offered by the Myrias parallel computers were of great value to this thesis work in realizing the potential of parallel processing concepts. The consulting skills of Dr. B. Minchau of the Myrias corporation have been particularly helpful to speed up the parallel programming phase of this thesis. I am also very grateful to the technical consultations from Mr. R. Heath and Mr. R. Morse of Applied Sciences in Medicine and Mr. N. Jantz of the Department of Electrical Engineering. Some of the test images used in this thesis work have also been provided by the Department of Applied Sciences in Medicine and I am extremely grateful for that. I thank with appreciation all the office staff of the Department of Electrical Engineering, who have been kind, helpful and encouraging throughout the program.

Last, but the most important of all is the support and encouragement during difficult times by family and friends. My special thanks are due, first to Mrs. Marianne Joerg-Koch and Mrs. Robin Robinson, whose cheerful and encouraging company have always been invaluable many times and in many ways to this thesis work. I also thank all my friends and colleagues and particularly those in Applied Sciences of Medicine, Electrical Engineering, Alberta Microelectronics Centre and ETH. The encouragement, understanding, support and cooperation from my husband and family members have always been the backbone for all my endeavours and this thesis marks their contribution. I dedicate this thesis to my father in the memory of his inspiring ideals and to my mother for her dedication in teaching me the values of life and to my husband for his support and encouragement to fulfill my goals.

TABLE OF CONTENTS

Chapter		Page
1.	**INTRODUCTION**	1
	1.1. The maximum entropy (*ME*) method - developments	1
	1.2. Image processing and machine vision	2
2.	**FOUNDATIONS OF THE MAXIMUM ENTROPY METHOD IN IMAGE PROCESSING**	4
	2.1. Application of entropy concepts to images	4
	2.1.1. Image restoration and reconstruction	5
	2.1.2. Image measurement model for restoration	6
	2.1.3. The *ME* restoration criterion	7
	2.2. Shannon's information measure basis	8
	2.2.1. The *ME* formalism	9
	2.2.2. Luminance frequency model of image formation	10
	2.3. The *ME* method using Bayes' theorem	11
	2.3.1. Bayes' theorem	12
	2.3.2. Bayes' theorem for images	12
	2.3.3. The *ME* method from Bayes' theorem	13
	2.3.4. Practical difficulties of the luminace frequency model	15
	2.4. Physical model basis	16
	2.4.1. Probability representation of a true image	17
	2.4.2. Statistical communication theory model of image formation	18
	2.4.3. Spectrum estimation basis for image reconstruction	19

2.4.4. The *logp* VS *plogp* forms of entropy 20
2.5. Frieden's statistical approach 24
2.6. The *ME* as a consistent variational principle 25
2.7. Investigation summary - I 31
2.8. Computational schemes 33
2.8.1. Newton - Raphson technique 34
2.8.2. Single constraint *ME* method 35
2.8.3. Optimization techniques 36
2.8.4. " Integral equation " approach 37
2.8.5. Entropy metric technique 37
2.8.6. Differential equation approach 38
2.9. Investigation summary - II 38

3. DEVELOPMENT OF A PARAMETER CONTROLLED RELAXATION SCHEME 42

3.1. Phase - I: Constrained iterative procedure 43
3.1.1. Specifications for computations 43
3.1.2. Solution equation 45
3.1.3. Interpretation 46
3.1.4. Initial condition 48
3.1.5. Intermediate variables and conditions 48
3.1.6. Error analysis and noise variance elimination 50
3.1.7. determination of σ_x^2 51
3.1.8. Termination condition(s) 53
3.1.9. Determination of λ_x and its difficulties 54
3.2. Phase-II: The constrained iterative procedure in a relaxation scheme 55

3.2.1. Relaxation of intermediate estimates with β 56
3.2.2. Influence of relaxation on the procedure 57
3.2.3. Practical form for λ_s 59
3.2.4. Significance of using β 60
3.3. Phase-III: Parameter control and convergence 63
3.3.1. Criteria for controlling λ and β 63
3.3.2. Controlling techniques 66
3.3.3. Convergence conditions in the procedure 66
3.4 Convergence mechanism of the parameter relaxation procedure 70

4. NUMERICAL AND FUNCTIONAL BEHAVIOUR OF THE RELAXATION SCHEME-TEST STUDIES 74
4.1. Test objectives and set-up 74
4.1.1. Choice of images and algorithms 74
4.1.2. The test algorithm 75
4.2. Phase -I: Numerical behaviour characteristics 76
4.2.1. Predicted Vs recorded behaviour 77
4.2.1.1. Constraint strength parameter λ 77
4.2.1.2. Relaxation parameter β 78
4.2.1.3. Scale parameter q 79
4.2.1.4. Entropy S 80
4.2.1.5. Convergence 80
4.2.2. Removal of upper-bound constraint: a test for convergence 81
4.2.3. Influence of K_λ on convergence speed 82
4.2.4. Error performance 83

4.3. Phase-II: Functional performance 84
4.3.1. Application perspective 84
4.3.2. Advantages of the *ME* method 85
4.3.3. Choice of source images 86
4.3.4. Generation of input images 86
4.3.5. Results of the *ME* estimation 88
4.3.6. Restoration in presence of noise-discussion 89
4.4. Performance Evaluation 90
4.4.1. The algebraic techniques of restoration 90
4.4.2. Comparative study 92
4.4.3. Mean square error evaluation 93
4.5. Observations for further investigations 95

5. MACH BAND BEHAVIOUR OBSERVATIONS -A SPECIAL TEST CASE STUDY 116

5.1. Test studies and discussions 116
5.1.1. Test results of the *ME* estimation 117
5.1.2. Mach bands 118
5.1.3. Mach band characteristics in biological visual systems 118
5.1.4. A comparison 119
5.1.5. Need for suitable parameter settings 119
5.1.6. The behaviour conditions of the algorithm for Mach bands 120
5.1.7. Test study inferences-summary 121
5.2. Analysis-stage I 121
5.2.1. Test class-I 122
5.2.2. Test class-II 123
5.2.3. Test class-III 124

5.3. Analysis-stage II 126

5.4. Interpretations 128

6. STUDIES ON THE MACH'S BIOLOGICAL MODEL OF VISUAL ENHANCEMENT 136

6.1. Studies on Mach's model 136

6.1.1. Mach's model-an overview 137

6.1.2. The underlying mechanism of enhancement 138

6.1.3. Boundary correction procedure 141

6.2. The type of Laplacian operator in the model 141

6.3. Implications 143

6.3.1. Unsharp - masking Vs Mach's model 144

6.3.2. Mach's model and *ME* edge enhancement 145

7. PARALLEL IMAGE CONVOLUTION AND *ME* IMAGE DECONVOLUTION-AN IMPLEMENTATION STUDY 151

7.1. The *ME* algorithms-a review 151

7.1.1. Dynamic programming techniques 152

7.1.2. Simulated annealing by stochastic relaxation 153

7.1.3. Parameter-controlled relaxation 154

7.2. Parallel Implementation 156

7.2.1. Myrias SPS-2 parallel computing system 157

7.2.2. The concepts of instruction and domain partitioning parallelism 158

7.2.3. Performance and efficiency measurements 162

7.3. Image convolution algorithm 163

7.3.1. Image convolution operation 164

7.3.2. Task specification strategy in the instruction parallelism approach 164

7.3.3. Performance studies with instruction parallelism 165

7.3.4. Performance studies with domain partitioning 167

7.4. Image deconvolution-the *ME* algorithm 168

7.4.1. Instruction parallelism approach 169

7.4.2. Domain partitioning approach-a discussion 170

8. CONCLUSIONS

8.1. The *ME* method-a general outlook 181

8.2. Computational scheme for the *ME* method 182

8.3. Application studies in Bayesian restoration 183

8.4. Links with biological models 184

8.5. Parallel implementation studies 185

BIBLIOGRAPHY 186

APPENDIX - A 198

APPENDIX - B 199

APPENDIX - C 200

APPENDIX - D 202

APPENDIX - E 204

APPENDIX - F 206

APPENDIX - G 207

APPENDIX - H 208

LIST OF FIGURES

Figure		Page
3	Mechanism of the parameter relaxation procedure	73
4.1	Algorithm for test studies	96
4.2	Inputs to the algorithm	97
4.3	Parameter control mechanism	98
4.4	Recorded parameter variations	98
	continued	99
	continued	100
	continued	101
4.5	Convergence test characteristics - removal of upper bound	101
	continued	102
	continued	103
	continued	104
4.6	K_λ Vs convergence speed	104
4.7	Intermediate error histogram variations	105
4.8	Source images and their frequency spectra	106
4.9	Transfer functions of the window filters	107
4.10	Generated input images and their frequency spectra	108
	continued	109
4.11	The degraded inputs and the *ME* restored images	110
	continued	111
4.12	The *ME* performance for reduced noise level input	112
4.13	Noise performance characteristics in the observed results	112

4.14	The *ME* Vs algebraic restoration	113
4.15	Performance of the *ME* method in restoring inputs of Fig. 4.11	114
4.16	Performance of the *ME* method in comparison with the standard methods	114
4.17	*ME* iterative improvements for the eye image in Fig. 4.11(d)	115
4.18	A test pattern and the observed *ME* estimate	115
5.1	Test features of the pattern in Fig. 4.18(a)	130
5.2	The 1-D intensity profile of the *ME* estimate of the test pattern	130
5.3	A family of input test patterns with varying intensity gradient widths	131
5.4	*ME* response to varying widths	132
5.5	*ME* response to illuminance changes	133
5.6	A sensitivity test pattern and the *ME* response	133
5.7	Characteristics of Mach bands from psycho-physical studies (schematic diagram)	134
5.8	Characteristics of Mach bands for varying widths from psycho-physical studies (schematic diagram)	134
5.9	Influence of β variations on the *ME* response to the test pattern in Fig. 5.1(a)	135
5.10	The iterative sequences of the test pattern in the *ME* estimation	135
6.1	Response of Mach's model to the test pattern 5.1(a)	147
6.2	Edge enhancement performance of the Mach's model	148
6.3	Response of Mach's model Vs *ME* estimation procedure to stimulus patterns with varying widths	149
6.4	Edge performance of the *ME* method-an example	150
7.1	Image convolution operation	172
7.2	Image convolution using instruction and data parallelism	172

7.3	Input image for the convolution algorithm	173
7.4	Speed-up and elapsed time performance characteristics	174
7.5	Output of the convolution algorithm	175
7.6	Time involvement of different computational phases in the *ME* image deconvolution	176
7.7	Image deconvolution algorithm using instruction parallelism	176
7.8	Speed-up characteristics for data space and estimate space calculations	177
7.9	Overall speed-up characteristics of the algorithm	177
7.10	Estimation of optimum number of processors from efficiency and elapsed time characteristics	178
7.14	Output of the image deconvolution algorithm	179
7.12	Behaviour characteristics of the relaxation procedure in a parallel algorithm	180

LIST OF SYMBOLS

CHAPTER - 2

ME	maximum entropy
n	true image size
m	measured image size
$\mathcal{N}_j$	number of luminance elements entering pixel j in an image formation process
$\mathcal{N}$	total number of luminance elements
n_j	number of photons in cell j
x_i	photon position i
N	number of experiment trials
f_i	true image pixel
$\mathbf{f} \equiv \{f_1, \ldots, f_n\} \equiv \hat{f}(i,j);\ i = 1, \ldots, \sqrt{n};\ j = 1, \ldots, \sqrt{n}$	true image distribution
$\mathbf{d} \equiv \{d_1, \ldots, d_m\} \equiv d(i,j);\ i = 1, \ldots, \sqrt{m};\ j = 1, \ldots, \sqrt{m}$	measured image distribution
$\hat{\mathbf{f}} \equiv \{\hat{f}_1, \ldots, \hat{f}_n\} \equiv \hat{f}(i,j);\ i = 1, \ldots, \sqrt{n};\ j = 1, \ldots, \sqrt{n}$	estimate of the true image distribution
$\mathbf{e} \equiv \{e_1, \ldots, e_m\}$	noise error distribution
$A \equiv \{Ajk, j = 1, \ldots, m;\ k = 1, \ldots, n\}$	measurement function
F_k, D_k	Discrete Fourier transforms of **f** and **d** respectively at frequency coordinate k = (k1,k2)
$p_i = p(\mathbf{x} = x_i)$	probability of an event x_i
$q_i = q(\mathbf{x} = x_i)$	prior probability of an event x_i

$\{p_1, \ldots\ldots p_n\}$	discrete probability distribution
$q \equiv \{q_1, \ldots\ldots q_n\}$	prior probability distribution
$f = \{f_1, \ldots\ldots, f_n\}$	true image under a luminance model of image formation
σ	standard deviation of noise error
ε	energy of a photon
z	number of degrees of freedom for photons
τ	temporal coherence time
λ	radiation wavelength
λ, β, μ	Lagrange parameters
δ, Δ	incremental qualifiers
$o, \Im$	information and coordinate transformation operator
$W(\mathbf{f})$	multiplicity factor for the distribution $\mathbf{f}$
$S(\mathbf{p})$	entropy of a probability distribution
$m(\mathbf{p})$	invariant measure function
Z	partition function
I	information
s	signal

Chapter-3

n	number of iterations
$\tilde{\tilde{\mathbf{f}}} \equiv \{\tilde{\tilde{f}}_1, \ldots, \tilde{\tilde{f}}_n\} \equiv \tilde{\tilde{f}}(i,j);\ i = 1, \ldots, \sqrt{n};$ $j = 1, \ldots, \sqrt{n}$	unbounded intermediate estimate
$\tilde{\mathbf{f}} \equiv \{\tilde{f}_1, \ldots, \tilde{f}_n\} \equiv \tilde{f}(i,j);\ i = 1, \ldots, \sqrt{n};$ $j = 1, \ldots, \sqrt{n}$	relaxed intermediate estimate
$\mathbf{p}$	path traced by an iterative algorithm
$\mathbf{S}$	a system

$\mathbf{s}_i$	a particular "normal" state of a system **S**
$\mathbf{s}_i$ or $\mathbf{s}_i$	states of the system **S** characterized by specific numerical behaviour
B	a scale for the numerical behaviour of the system
$\varepsilon, \xi, \zeta, e$	different error terms
$\sigma_\xi^2, \sigma_\zeta^2, \sigma^2$	different variance terms
$\overline{\lambda}$	recursive form of λ
γ	convergence measure
β	relaxation parameter
q	scale parameter
$g(x)$	iterative function
$g'(x)$	first derivative of the iterative function
$\| \|$	modulus operator

Chapter-4

F	Fourier transform operator
$\mathbf{F}^{-1}$	inverse Fourier transform operator
NMSE	normalized mean square error
SNR	signal to noise ratio

Chapter-6

um	unsharp masking
r_p	neural activity at point p
B_p	brightness at point p
I_p	intensity at point p
L	Laplacian operator

$O(f(n)) \equiv \{g(n) \mid \exists\, c, n_0 \in N \; |g(n)| \leq cf(n) \; \forall \; n > n0\}$

	time complexity: order at most f(n)
N1	image dimension in X - direction
N2	image dimension in Y - direction
M1	Window filter dimension in X - direction
M2	Window filter dimension in Y - direction
J(x)	cost or an objective function defined over x.
P	a general program
P_s	an execution program
τ	time interval
t_o	initiation time of an execution
$\{I_j\}$	Instruction set of a program
D	a data set pertaining to the instructions of a program
D	data set pertaining to a particular set of instructions
D	input data domain
d	partitioned domain
η	algorithm efficiency
p	number of processing elements
S(p)	speed-up ratio using p processors
t(p)	cpu(central processing unit) execution time
e_t	cpu elapsed time
T	a *pardo* task
T_c	a child task of the *pardo*.

General mathematical symbols

$\mid$	conditional implication A|B (A given B)
$\equiv$	equivalence
$\varnothing$	an empty set
$\cup$	set union $A \cup B = \{x \mid x \in \text{or } x \in B\}$
$\cap$	set instruction $A \cap B = \{x \mid x \in \text{and } x \in B\}$
$\subseteq$	set inclusion
$\in$	set membership
$\notin$	set nonmembership

CHAPTER 1

INTRODUCTION

The origin of the entropy based methods dates back to the era of Bernoulli, Laplace, Bayes and Gibbs, when attempts were made to solve problems of inversion by the principle of insufficient reason and plausible reasoning approaches [1], [2]. The concepts were found useful in deriving equilibrium conditions in statistical mechanics and thermodynamics. The turning point occurred when C.E. Shannon introduced the concept of entropy as a unique measure of uncertainty represented by a probability distribution in transmitting messages over communication channels [3]. Contributions were made in the analysis of channel capacity and efficiency of codes using this measure. The second turning point evolved when E.T. Jaynes introduced the concept of *maximizing the entropy* for general scientific inferences, which is the foundation for most of the current entropy methods [4], [5]. His concept is based on Shannon's interpretation of entropy. A related approach in information theory is referred to as minimum directed divergence, as minimum discrimination information, or as cross-entropy minimization [6]. Although the history of developments is quite long for the methods, their usefulness to practical applications was not realized until the introduction of the entropy maximization concept.

1.1. The maximum entropy *(ME)* method - developments

The method based on the entropy maximization concept in its basic form involves maximizing the entropy subject to the input data as constraints. From now on, the term "entropy methods" will be replaced by the term "*ME* method". So far, the method has been successfully applied to two major application areas, namely, powerspectrum estimation and image restoration and reconstruction. The application of the method to power spectrum estimation by J. P. Burg, accounted for the side-lobe discrepancies associated with the

linear techniques [7]. It was in this work that the consequences of the assumptions used in linear techniques were realized for the first time. The method is relatively well-established in spectrum estimation [8], [9], [10] compared to image processing. The usefuleness of the *ME* method in image restoration was first established by R. Frieden. He showed that a positive restoring formula logically arises from a statistical communication theory model of image formation, and the most likely object implied by the image data obeys the principle of *ME* [11]. In spectrum estimation, as well as image restoration, the improved resolution of the results are commonly attributed to the 'inherent positivity' and the smoothness criteria that are fulfilled by the *ME* estimates. Following Frieden's work, the *ME* method supplemented with efficient noise reduction schemes was applied to many radio astronomy applications of image reconstruction [12], [13], [14], [15]. In its subsequent applications in image enhancement and restoration, the method has been reported to yield outputs that are reliable and relatively free from artifacts [16], [17], [18], [19], [20]. Superior results have been reported so far in areas such as computed tomography, X-ray, diffraction and positron emission tomography, radiography, radio astronomy, optical deconvolution, crystallographic diffraction imaging, NMR spectroscopy and forensic imaging. The applications are scattered throughout the literature and a collection of it appears in some of the references [16], [21]. The various developments have also led to the idea of the entropy maximization as the only consistent variational technique [22], [23]. The current applications attempt to make use of the prior knowledge incorporation features of the *ME* method, particularly in the fields of expert systems and neural networks [24], [25], [26], [27].

1.2. Image processing and machine vision

In image processing, techniques are broadly classified as image restoration, reconstruction, enhancement, coding and segmentation [28], [29], [30]. The need for these techniques arises in many fields such as medical diagnostics, atmospheric remote

sensing, radio astronomy and microscopy. One of the related areas of recent interest is machine vision or computational vision, where efforts are focused in the direction of image analysis and interpretation, leading to the machine recognition of an object [31], [32]. Recovering 3-D properties of surfaces from the 2-D images, extraction of features, classification a.e some of the areas of research interest in image analysis. These applications invariably involve manipulation and extraction of information from 2-D images such as optical flow, edge detection, image restoration, and interpolation. Most of the image processing and computational vision operations have been recognized to be members of the family of inverse problems [33], [34]. These problems are ill-conditioned and ill-posed and a satisfactory general solution has not yet been established due to characterization difficulties. In solving inverse problems, the use of prior knowledge or information is considered important. But there exists conceptual and modelling difficulties for images in general at computational levels and also at levels of incorporating prior information or combining evidences [35], [31].

The natural flexibility of the *ME* method in incorporating prior information or extra knowledge appears to be well-suited to many problems of image processing and machine vision. But the method in its present form has serious progress limitations. In particular, the criticisms regarding the foundations of the method have a dominant influence. The theoretical foundations of the method are often condemned as philosophical assertions. While the criticisms on the theoretical issues are only at the level of disputes, the practical difficulties of the method are more fundamental and realistic. Due to the computational difficulties and complexities, the methods are considered impractical specifically for image processing applications. As a matter of fact, at the computational level, the method has not advanced beyond its current application areas of image restoration and reconstruction. Even in these areas, there lies the fundamental difficulty of obtaining a solution for the method. The method apparently needs further investigation and development in order to achieve its full range of potential application.

CHAPTER 2

FOUNDATIONS OF THE MAXIMUM ENTROPY METHOD IN IMAGE PROCESSING

The *ME* method is well known in the area of time-series spectral estimation to estimate a power spectrum of a stationary random process from a finite but incomplete number of auto correlation coefficients. In image restoration or reconstruction, the *ME* method is employed to estimate a true image from its measured samples that are finite, distorted and incomplete. Unlike the time-series case, images are 2-D spatial power patterns often resulting from a quantum statistical process. Extension of entropy concepts to images have introduced different interpretations and view points in the literature. This chapter investigates the basis of the *ME* method for images by closely following its various development stages.

2.1. APPLICATION OF ENTROPY CONCEPTS TO IMAGES

A number of physical problems are modeled using an expression of the form:

$$O(n_k) = D[I(n_i)] \; ; \; i = 1, \cdots\cdots n \qquad k = 1, \cdots\cdots m; \; m \leq n \tag{2.1}$$

where D[.] is a distortion or a transformation operator that acts on the multi-dimensional (including one-dimensional) input sequence $I(n_i)$ to produce an output sequence $O(n_k)$. In many real applications, the $O(n_k)$ are the *measurements* of continuous physical quantities and the problem is to determine the input I from O, which is called an *inverse problem* [36], [37]. Deconvolution, noise removal and signal extrapolation are some of the common examples of inverse problems. The *ME* method is reported to be a solution method for inverse problems in general [38], but has not been firmly established beyond spectral

estimation of 1-D time signals. Although the basic concepts are similar, extension of the method to 2-D signals in general, and images in particular, is not well understood. In this chapter, attention is focused on 2-D images. Considering the example of image restoration and reconstruction, which includes the problems of deconvolution, noise removal and signal extrapolation, the *ME* method is discussed in the sections that follow.

2.1.1. Image restoration and reconstruction

True images are real, non-negative and bounded spatial power patterns of a continuous *object* field. The patterns are often represented as 2-D functions f(x,y), resulting either from a deterministic or a random process [28]. The *physically measured* quantities are the spatial samples of f(x,y), characterized by the measurement medium. In situations where a direct measurement is not possible, some function related to f(x,y) will be measured. In any case, there is a practical necessity to obtain f(x,y) from the measured image. The processing techniques employed for the fundamental purpose of obtaining a continuous true image in a quantitative manner are classified as image restoration and reconstruction techniques in image processing [28], [29], [30], [39]. Although the basic objective is the same, the two techniques differ in terms of the processing conditions. Specifically, inputs for restoration techniques are the degraded measurements involving physical processes such as atmospheric effects, lens distortion, channel and sensor noise, echoes, reflections. Reconstruction, on the other hand, applies to inputs with sampling problems. However, in the *ME* literature, as we will see in the following section, the term 'reconstruction' is often referred to as a general term meaning restoration as well as interpolation. The ambuiguity is not a matter of major concern from applications point of view, since both techniques follow the same mathematical representation for using the *ME* method. However, in discussing the foundations of the method, the distinctions appear to be important. At this point, let us choose to discuss the subject matter from the view point of restoration since the concept is more involved.

Image restoration in general involves five stages of development: an image measurement model, a restoration criterion, a computational procedure, an algorithm and its implementation. The measurement model is an imaging model that tells us how the recorded data are related to the true image data. The restoration criterion specifies the conditions the solution needs to satisfy in order to restore the true image. Stated simply, it specifies what restoration means. The computational procedure defines the solution methodology for the restoration criterion and the algorithm is a means to carry out the procedure on a given implementation scheme [40], [41]. In this chapter, we are concerned with the first two stages of development, with more emphasis on the restoration criterion as applied to the *ME* method.

2.1.2. Image measurement model for restoration

Consider a true image of an object represented by a set of finite samples, denoted as $\mathbf{f} = \{f_1,....,f_n\}$ and measured as $\mathbf{d} = \{d_1,....,d_m\}$ with $m \leq n$. Strictly speaking, a general measurement model follows the form in (1) with a general class of degradation processes. However, for discussion purposes, let us consider a widely followed restoration model [39]. Under this model, the measured image datum d_j is written as

$$d_j = \sum_{k=1}^{n} f_k A_{jk} + e_j \quad j = 1,\ldots\ldots,m \quad m \leq n . \qquad (2.2)$$

1-D notation is followed for convenience. The term e_j refers to the noise error (here assumed to be additive and zero mean Gaussian) present in the datum d_j. A is a 2-D function $A = \{A_{jk} ; j = 1,.....,m ; k = 1,.....n\}$ that relates **d** and **f** in a measurement system and is assumed to be known. For convolved data A is commonly a linear, space-invariant, point spread function of the measuring system. In using (2.2), it is of fundamental importance to note that the model exclusively refers to monochrome image

class. The restoration approach differs significantly for the case of multi-spectral image class. The importance of this distinction is brought out in [42]. In this work, the restoration analysis and the schemes refer to monochrome images. The objective of restoration is to find the true image $\mathbf{f}$, or the most probable image $\hat{\mathbf{f}}$, from the measurements $\mathbf{d}$ given by (2.2).

2.1.3. The *ME* restoration criterion

The problem of finding $\hat{\mathbf{f}}$ from (2.2) is an ill-conditioned and ill-posed inverse problem because, in practice, the data are often incomplete ($m < n$) and noisy with a singular A. The under-constrained and the corrupted nature of the problem introduces an infinite number of possible solutions, called *feasible estimates* that could have given rise to the measured data. As stated in [43], "this many-to-one mapping from an object to image is an intrinsic property of any real measuring system that has a finite aperture or bandwidth and hence a point-spread function of non-zero width". Finding an estimate of $\mathbf{f}$ means choosing a particular estimate from the set of feasible estimates based on a criterion that is 'best' in some sense. Two types of restoration techniques are used in practice [39], [44]: algebraic and statistical. The algebraic restoration techniques make use of approximation techniques and filter transfer function design techniques. The statistically based restoration techniques, on the other hand, make use of restoration criteria based on the statistical formulation priciples of an image. The *ME* method is a statistical approach based on a photon random emission models of image formation. In the restoration process the method follows the criterion of assuming the least about the unknown measurements, while allowing the incorporation of any known information about the true image. How the criterion is derived from statistical principles is an interesting issue underlying the *ME* method. A more interesting fact is the multiple basis for the criterion, which means that the criterion is derived as a consequence of different view points and interpretations. The existence of multiple basis is a major source of confusion in the literature. Also, because of

extensive arguments and criticisms, the basis of the method rarely appears in clear terms and is often subject to many misinterpretations in the literature. The different view points for the method arise from using:

i) Shannon's information measure of uncertainty,

ii) the Bayesian theorem,

iii) physical models,

iv) Frieden's statistical approach,

v) consistent variational principles.

The different view points are explained in the following sections with the intent of clarifying the underlying concepts. Quotes and italic fonts are employed to emphasize unique and important terms.

2.2. SHANNON'S INFORMATION MEASURE BASIS

Let us start with the understanding of the *ME* method as proposed originally by E.T. Jaynes [4], [5], [38]. Let us suppose that a real variable **x** can take any one of the discrete values $\{x_1,.....x_N\}$ and the information about **x** is available as constraint(s) of form (2.2). According to Jaynes, it is possible to *assign* a probability distribution $\{p_1,.....p_N\}$ (p_i is written as shorter notation for $p(\mathbf{x} = x_i)$) in order to *represent* the given partial *information* about **x**. The probability distribution is obtained by maximizing its entropy subject to the available information as constraints. The entropy of the distribution is given by Shannon's measure of uncertainty in a *discrete* probability distribution:

$$S(p) = -K\sum_{i=1}^{N} p_i \log p_i \qquad (2.3)$$

K is a constant that allows for an arbitrary logarithmic base. Jaynes also discusses the entropy of the continuous case [5],[45] as:

$$S(p,m) = -K\int p(x)\log\left[\frac{p(x)}{m(x)}\right]dx \qquad (2.4)$$

He calls m(x) an 'invariant measure' function, the reasons for which will become clear in section 2.5. For all practical purposes, let us consider the form (2.3). The basis of the method as discussed in [2], rests upon the consistency properties of (2.3) as uncertainty measure and hence upon Shannon's derivations.

2.2.1. The *ME* formalism

Putting aside, for now, the image measurement model (2.2), the *general* formalism of the *ME* method proceeds as follows [38]. Let us consider the 'data', a set of numbers $\{A_1',.....A_m'\}$ resulting from m known functions $\{A_1(x):.....A_m(x)$. In the *ME* approach, the data are the values a probability distribution $\{p_1,.....p_n\}$ is required to 'predict'. To 'fit' the distribution with the 'measured data', m simultaneous constraints are imposed:

$$\sum_{i=1}^{n} p_i A_k(x_i) = A_k \qquad (2.5)$$

If the 'predicted' distribution $\{p_i, ...p_n\}$ is substituted in the left hand side term of (5), it should give rise to the m numbers on the right hand side. The distribution is the *ME* distribution found by maximizing (2.3) subject to (2.5) and is given by

$$p_i = \frac{1}{Z(\lambda_1,.....,\lambda_m)}\exp[-\lambda_1 A_1(x_i) - ..... - \lambda_m A_m(x_i)] \qquad (2.6)$$

where

$$Z(\lambda_1,....., \lambda_m) = \sum_{i=1}^{n}\exp[-\lambda_1 A_1(x_i) - ... - \lambda_m A_m(x_i)] \qquad (2.6.1)$$

with λ's determined by solving

$$A_k = -\frac{\partial(\log Z)}{\partial\lambda_k} \qquad (2.6.2)$$

The application of the formalism to images is discussed below.

2.2.2. Luminance frequency model of image formation

In extending the general *ME* formalism to practical applications, means are necessary to define the application in terms of probability events. Jaynes understands image 'reconstruction' (includes restoration as well) as the process of distributing $\mathcal{N}$ equal elements of luminance distributed over n pixels, labeled $\{x_1, \ldots\ldots, x_n\}$ to form an image with $\mathcal{N}_i$ elements of luminance at pixel i [46]. This means a particular result x_i is generated with 'frequency' $\mathcal{N}_i / \mathcal{N}$ and the 'scene' (image represented in terms of elements of luminance) $f = \{f_1, \ldots\ldots f_n\}$ represents the true image distribution **f** (note: change in font type). The number of ways in which the true image distribution could have been generated is the *multiplicity factor* $W(f)$ calculated as

$$W(f) = \frac{\mathcal{N}!}{\prod_{j=1}^{n} \mathcal{N}_j !} \tag{2.7}$$

With large $\mathcal{N}$, Stirling's approximation of $\log W(f)$ gives rise to the form

$$\log W(f) \cong \mathcal{N} S(f) \tag{2.7.1}$$

where

$$S(f) = -\sum_{i=1}^{n} f_i \log f_i \tag{2.7.2}$$

From (2.7) and (2.7.1) we understand that the 'entropy' (2.7.2) contains the multiplicity information (the number of possible distributions) pertaining to any true image **f** that falls under the luminance model. In the present context, we are interested in finding **f** that is hidden in the measured image data given by (2.2). The measured image data, as R. Frieden

puts it are 'blind' to the order $\{x_i\}$ in which the sequencing would have taken place in forming the image. What we know is a set of numbers $\{d_1,.....d_m\}$ as the average of m quantities:

$$\sum_{i=1}^{n} N_i A_k(x_i) = d_k \; ; 1 \le k \le m \qquad (2.8)$$

(it may be noted that the noise terms have been omitted in writing (2.8)). The most likely frequency distribution $\{\hat{f}_1,.....\hat{f}_n\}$ according to Jaynes is the one that maximizes the multiplicity $W(f)$ subject to the constraints (2.8). The solution is given by (2.6) with p_i replaced by $\hat{f}_i$. Two aspects of the approach need further considerations: The incorporation of noise reduction capabilities and the practical significance of N in restoration. Using the *ME* prior information in Bayes' theorem, Jaynes suggests a possible way of incorporating noise reduction capabilities into the *ME* method [38]. The second aspect is more of the shortcoming of the model as we will see shortly.

2.3. THE *ME* METHOD USING BAYES' THEOREM

The basis of the *ME* method using Bayes' theorem, discussed in this section pertains to the specific context of images, where the correlations between the adjacent pixel luminances are not a major concern. It is important to realize this fact because, often in the literature, the Bayes' theorem basis of the method is reported as a general statement, while the connections between the two positions is still a research issue. In this section, the derivation of the *ME* method is considered with the intent to clarify the approach in the light of the Bayes' theorem. Besides, the approach has significant relevance to the subject matter of the subsequent chapters.

2.3.1. Bayes' theorem

Bayes' theorem has a long history, suggesting the idea of taking the prior distribution and the likelihood into account to obtain solutions for inverse problems [1]. The foundations of Bayes' theorem are built on the rules of inductive logic among a set of propositions, making use of the consistency requirements [2]. Applying these requirements to a specific proposition p(AB|C) (i.e. probability that propositions A and B are true given C) leads to the relation

$$p(A|BC) = \frac{p(B|AC)\,p(A|C)}{p(B|C)} \qquad (2.9)$$

which is well known as Bayes' theorem. Although the Bayes' theorem and the maximum entropy methods exhibit similarities on the grounds of consistency, the context in which they apply differ. In (2.9), the constituting propositions are *probabilities*, and the *calculated* result is a *single* quantity. On the other hand, the problem of solving for $\hat{\mathbf{f}}$ in (2.2) pertains to an *inference* problem with *distributions*. A direct application of (2.9) is therefore not possible.

2.3.2. Bayes' theorem for images

It is however possible to obtain probability distributions using statistical processes and models that allow (2.9) to be written as

$$p(\mathbf{f} \mid \mathbf{d}, \sigma, I) = p(\mathbf{f} \mid \sigma, I)\,\frac{p(\mathbf{d} \mid \mathbf{f}\,\sigma I)}{p(\mathbf{d} \mid \sigma\, I)} \qquad (2.10)$$

where I is the prior information. $p(\mathbf{f} \mid \sigma, I)$ [$= p(\mathbf{f} \mid I)$] and $p(\mathbf{d} \mid \mathbf{f}\sigma I)$ are called *a priori* probability distribution and *likelihood* respectively. The denominator term is independent of $\mathbf{f}$ and hence is treated as a normalizing constant and has no specific name. The result is called the *a posteriori* probability distribution. Maximizing the a posteriori probability distribution is one of the popular image estimation methods, called maximum a posteriori methods. One of the key features of the approach is the prior probability distribution. The

various possibilities of obtaining this term can be found in the literature [47]. One such possibility, as E.T. Jaynes points out, is the use of entropy.

2.3.3. The *ME* method from Bayes' theorem

In one of his works [38], E.T. Jaynes shows a possible way of estimating the prior probability distribution with the Bayes' theorem. Based on the luminance frequency model of image formation, (2.2) can be written as

$$d_k = \sum_{i=1}^{n} A_{ki}\mathcal{N}_i + e_k \; ; \; 1 \le k \le m < n \tag{2.11}$$

where $\mathcal{N}_i = \mathcal{N}f_i$. The Gaussian noise terms e_k follow a probability distribution function

$$p(e_k \mid \sigma) \propto \frac{1}{\sigma} \exp\left(-\frac{e_k^2}{2\sigma^2}\right) , \; 1 \le k \le m \tag{2.12}$$

where σ is the standard deviation of the noise terms and $\propto$ denotes proportionality. Now let us consider the problem of finding $\hat{\mathbf{f}}$ in (2.2). As we already know, it is an ill-posed problem, with infinite number of feasible **f**-distributions. In the present context, we are trying to translate this conceptual description into an equivalent mathematical representation. According to Jaynes, the multiplicity information provides a possible medium. Based on the luminance frequency model, his ideas proceed in the following manner: there are $n^{\mathcal{N}}$ conceivable ways and a particular distribution $f = \{f_1, \ldots\ldots f_n\}$ can be realised in $W(f) = \exp\{\mathcal{N}S(f)\}$ ways. With the understanding of the 'prior scene' as the 'infinite number of *f*-distributions', Jaynes establishes the medium as

$$p(f \mid I) = \frac{1}{n^{\mathcal{N}}} \exp\left\{\mathcal{N}(S(f))\right\} . \tag{2.13}$$

Based on the justifications in [45], he calls (2.13) a prior probability distribution. This distribution sets up only the medium for the representation. When we add the meaning of

'feasible' to the possible f-distributions, we arrive at a final representation of the measurements given by (2.2). The likelihood distribution fulfills this requirement. It is calculated as follows: Given that the true scene is f, the probability density that we shall obtain the data $\mathbf{d} = \{d_1,.....d_m\}$ can be written directly from (2.2) and is just the probability that the noise terms e_k will make up the difference. It is given by

$$p(\mathbf{d} \mid f,\sigma,I) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-Q(\mathbf{d},f)}{\sigma^2}\right) \tag{2.14}$$

where

$$Q(\mathbf{d},f) = \frac{1}{2}\sum_{j=1}^{m}\left|d_j - \mathcal{N}\sum_{i=1}^{n} A_{ji} f_i\right|^2 . \tag{2.14.1}$$

Substituting (2.14) and (2.13) in (2.10) gives rise to the *a posteriori* probability distribution:

$$p(f \mid \mathbf{d}, \sigma, I) \propto \exp\left[\mathcal{N}\{S(f) - w\,Q(f)\}\right] \tag{2.15}$$

where

$$w = \frac{1}{\mathcal{N}\sigma^2} . \tag{2.15.1}$$

From (2.15) we see that maximising the *a posteriori* probability distribution is equivalent to maximising the function

$$J(f) = \mathcal{N}S(f) - \frac{Q(\mathbf{d},f)}{2\sigma^2} \tag{2.16}$$

which is the *ME* method in Bayes' theorem approach. The *ME* distribution obtained by solving $\partial J/\partial f_j = 0$ for a particular pixel j obeys the relationship

$$-\mathcal{N}(1+\log f_j) - \frac{1}{\sigma^2}\frac{\partial Q(\mathbf{d},f)}{\partial f_j} = 0 . \tag{2.17}$$

The solution is implicit in (2.17) and numerical procedures are necessary to find the estimate $\hat{\mathbf{f}}$. Before proceeding with the solution, it is necessary to understand and resolve the ambuiguities related to $\mathcal{N}$.

2.3.4. Practical difficulties of the luminance frequency model

The luminance model explained above is reported to have significant practical difficulties [23], [48]. Let us consider the *ME* solution given by (2.17). The first term comes from the prior probability assignment and the luminance frequency model with $\mathcal{N}$ denoting the number of luminance elements. As $\mathcal{N}$ increases, the first term of the equation dominates giving rise to a uniform grey image, which is unrealistic. Therefore, for all practical purposes, it is argued that $\mathcal{N}$ cannot be infinite. Although the arguments appear reasonable on practical grounds, it may be noted that Jaynes approach is to some extent misinterpreted. Jaynes identifies $\mathcal{N}_i/\mathcal{N}$ as 'frequency' and not as 'probability' but the controversies are raised by using it as a probability term in (2.3). In a different context, Jaynes makes it clear that as $\mathcal{N}$ increases in (2.17), σ decreases (σ due to the variability of the scenes) and hence it is $\mathcal{N}\sigma^2$ or w that determines the solution conditions and not $\mathcal{N}$ alone [38]. But still there is an element of ambiguity in interpreting σas the spread due to variability of scenes, because, σ as we noted in the derivations, is the spread due to noise. There are many other ambiguities for $\mathcal{N}$ with respect to its usage and physical meaning [23]. From (2.16), we see that for any given $\mathcal{N}$, at the maximum of J, f is expected to obey the relation

$$2\mathcal{N}\frac{\partial S(f)}{\partial f_i} = \frac{\partial Q(\mathbf{d},f)}{\partial f_i}. \tag{2.18}$$

$\mathcal{N}$ is then given by

$$\mathcal{N} = \frac{\dfrac{\partial Q(\mathbf{d},f)}{\partial f_i}}{2\dfrac{\partial S(f)}{\partial f_i}}. \tag{2.19}$$

(2.19) shows that $\mathcal{N}$ depends upon the function Q, which means that it needs to be chosen *a posteriori* and is not consistent with using $\exp(\mathcal{N}S)$ as *a priori* probability distribution. Other questions raised in the literature are: should $\mathcal{N}$ be electrons per unit cell, atoms or photons?; What type of physical analysis is appropriate to determine $\mathcal{N}$?. The consequences of using $\mathcal{N}$ are also demonstrated experimentally in many specific applications. It should however be noted that R. Frieden, in the very first application of the *ME* method to images, eliminated the practical difficulties associated with $\mathcal{N}$ [11] using the physical basis of true image formation.

2.4. PHYSICAL MODEL BASIS

For many types of images, intensities result from the emission of discrete particles called photons from objects, specifically incoherent optical objects [28], [39]. Let us consider such an object, partitioned into n pixels, denoted as $\{x_1,.....x_n\}$ each of area ΔA. Let the i^{th} pixel of the image be identified with an average intensity (brightness) $f(x_i)$. In terms of the number of photons and their energies, the i^{th} pixel intensity is given by

$$f(x_i) = \frac{\varepsilon}{\Delta A} r(x_i) \qquad (2.20)$$

where ε is the photon energy and r_i is the number of photons emitted from the i^{th} pixel. Since the photon emission process is random, an image is considered to be a result of a statistical process, often in a quantum sense. In this random emission model of image formation, the probability that a photon was emitted from the pixel i is given by the normalized pixel intensity itself. A more detailed physical explanation in the context of

optical objects follows [49].

2.4.1. Probability representation of a true image

A photon originates from within an area Δx of a true image plane and follows an *unknown* path through a lens imaging system and strikes the image plane forming the image intensity profile $\mathbf{d}(\mathbf{x})$ (1-D notation followed for discussion convenience). The photons are assumed to have equal wavelength λ (a monochromatic image formation). In this situation, we are interested in answering the question: what is the probability $p(\mathbf{x} = x_i)$ that a photon will be emitted from the i^{th} cell of the true image plane?. The event is the emission of a photon from pixel i. The number of photons r_i emitted from cell i determine the number of events that took place to form pixel i, and is proportional to the energy flux $f(x_i)\,\Delta x$. The total number of emitted photons is proportional to the total energy flux, $\sum f(x_i)\,\Delta x$. Since the total number of photons is very large for the majority of optical cases, by the law of large numbers, the true image can be represented as $\{p(x_1),.....,p(x_n)\}$, or in short $\{p_1,.....,p_n\}$ where

$$p(\mathbf{x} = x_i) = p_i = \frac{r_i}{\sum_{i=1}^{n} r_i} = \frac{f_i}{\sum_{i=1}^{n} f_i} \qquad (2.21)$$

With $\sum_i f_i$ normalized to 1, the probability of the position for a photon in the true image plane equals the true image intensity distribution itself. Besides its physical basis, the form (2.21) has been reported to be free from the technical difficulties encountered with $\mathbf{N}$ [18], [47]. For example, it shows consistency of a uniform image with the 'equal weight' prior probability assumption of $1/n$ when used with the *ME* method. The denominator of (2.21) is the total light intensity and has an interpretation of a normalizing factor. The nomalization

does not imply that the probabilities are independent of the total intensity of light. This is because the numerator term f_i varies with the illumination conditions due to the changes in reflectivity.

2.4.2. Statistical communication theory model of image formation

With the probabilistic interpretation of images given by (2.21), R. Frieden extends Jayne's model for images and calls the model a statistical communication theory model of image formation [11]. When an object is subdivided into J equal-sized, elemental cells, each of area ΔA, it means that $J\Delta A = n$. In this context, ΔA is the 'spatial resolution' limit. Along the same line of reasoning, the intensity values can be expressed using a limit $\pm\Delta f$, which R. Frieden calls the 'uncertainty' limit. Perhaps both Frieden's as well as Jayne's models are better explained using a simple term 'quantization'. In Frieden's approach, the spatial as well as the intensity values are quantized with levels ΔA and Δf. Using this model, true images are represented by

$$f_j = N_j \Delta f \text{ for } j = 1,\ldots n, \tag{2.22}$$

where N_j is a dimensionless number with Δf containing the unit of radiance. We say there are $N_j \Delta f$ 'object units' in the pixel j. The total number of object units consituting a given image are

$$\sum_{j=1}^{n} f_j = N \Delta f. \tag{2.23}$$

The probability that any one object unit being located in cell j is p_j and is given by (2.21). Let us assume that the occurrence of one object unit in a cell j does not affect the possible location of any other object unit. That is the object units are statistically independent *within each cell* (uniform prior) and also from cell to cell. The number of ways W(**f**) that an

object can occur is the same as Jayne's model given by (2.7) but with $W(f)$ replaced by $W(\mathbf{f})$. Thus we see that, the missing element in Jaynes luminance model is the unit of radiance, which comes as a result of quantizing the true image distribution in Frieden's approach. Taking logarithm and applying Stirling's approximation for (2.7), Frieden writes the final expression as

$$\mathbf{N}S(\mathbf{f}) = -\frac{I}{\Delta f}\sum_{j=1}^{n} f_j \log f_j \qquad (2.24)$$

He calls the resulting *flogf* form of entropy as the *spatial entropy* of the true image. On mathematical grounds, it is however not clear how Frieden arrives at the form (2.24). Taking the logarithm and applying Stirling's approximation to (2.7) using Frieden's model we get

$$\log W(\mathbf{f}) \equiv \mathbf{N}S(\mathbf{f}) \qquad (2.25)$$

where $S(\mathbf{f})$ is given by

$$S(\mathbf{f}) = -\sum_{j=1}^{n} \frac{f_j}{\Delta f} \log\left(\frac{f_j}{\Delta f}\right) \qquad (2.25.1)$$

Substituting for Δf from (2.23)

$$S(\mathbf{f}) = -\mathbf{N}\sum_{j=1}^{n} \frac{f_j}{\sum_{k=1}^{n} f_k} \log\left(\frac{f_j}{\sum_{k=1}^{n} f_k}\right) \qquad (2.25.2)$$

Unlike Jayne's model, $\mathbf{N}$ is a fixed constant in Frieden's approach. The difference between (2.25.2) and (2.7.2) is the model, which allows S to be written in terms of probabilities rather than frequencies. All the results explained under Bayes' theorem approach are therefore applicable with the frequencies replaced by pixel intensities.

2.4.3. Spectrum estimation basis for image reconstruction

Wernecke and Addario in their radio astronomical image 'reconstruction' (does not

refer to restoration) application observes a different situation [12], [50]. In their application, an incoherent *radio source* gives rise to a random electric field, whose spatial autocorrelation function is sampled by radio interferometers. The unknown radio brightness distribution and the available interferometer measurements are related to each other by Fourier transformation. The objective is to reconstruct the brightness distribution from incomplete auto correlation measurements. They consider the problem as the power spectral analysis of the electric field. Their arguments are not based on any physical model, but their application is based on the fact that entropy rate $S(f)$ of a stationary, band-limited random process is related to its power spectrum through the relation

$$S(f) = \int_{-f_c}^{+f_c} \log\omega(f)df \qquad (2.26)$$

where f_c is the cut-off frequency. Thus they introduce the *logf* form of entropy for image reconstruction. According to them, the use of the *plogp* probability form in (2.25.2) is not suitable for radio astronomical images since the measurements do not have an optical basis of random emission of photons. They also raise an objection to the spatial entropy of images introduced by Frieden because maximizing the *flogf* form is not equivalent to maximizing the *plogp* form unless the total intensity is assumed exactly known, which is unrealistic. Irrespective of these differences, their work introduced a new concept in the *ME* literature. The concept suggests that the choice of entropy form for a particular application depends upon the physics of the measurement process. The concept also implies that the information theoretic *plogp* form is not justified adequately for all types of applications.

2.4.4. The *logp* Vs *plogp* forms of entropy

In an attempt to clarify the confusion, the *logp* and *plogp* forms were analysed in the context of 'image restoration' by Kikuchi and Soffer [43]. Their terminology of

'restoration' includes reconstruction problems as well, but their analysis is based on the photon emission model of image formation. Also, they refer to the *logp* and *plogp* forms as *logB* and *-BlogB* forms because, with the photon emission model, p_j is considered as the normalized brightness (intensity) B. They discuss the conditions under which the two forms should be used. Their analysis starts with the physical interpretation of an image as a 2-D spatial power pattern of an object formed by photon sources in the farfield. The emitted or the reflected photons from the object are received by the true image image plane **f** in an *observation time frame* of t. The fundamental question is the same: where (which position i of the object) does the photon come from?. The answer is not straightforward because, the time as well as the spatial sequence of photon positions during the time t in forming the true image **f** is unknown. They find an answer by investigating the underlying *physical causes* of uncertainties for the photon cell positions. They find that the causes are mainly the exposure time t, the temporal coherence time $\tau = 1/\Delta\nu$ for the radiation of bandwidth $\Delta\nu$, the area A of each detector aperture in the image plane and the coherence area σ subtended by the resolution cell at the image plane. These physical factors influence the process of image formation by altering the *states of photons* in each cell. Considering the *degrees of freedom* z (number of possible states) as the parameter for characterizing the image formation process, they derive z in terms of the various influencing factors mentioned above. They derive it to be

$$z \equiv z_t z_s = \left(\frac{t}{\tau}\right)\left(\frac{A}{\sigma}\right) \qquad (2.27)$$

The subscripts z_t and z_s stand for temporal and spatial degrees of freedom respectively. The number of photons n_j of cell j are distributed over the z degrees of freedom within the bandwidth $\Delta\nu$ with the condition that more than one photon can occupy in any one degree of freedom. To find the number of possible photon arrangements, they make use of the Bose-Einstein expression and write the number as

$$q_j(n_j) = \frac{(n_j + z-1)!}{n_j!\,(z-1)!} \qquad (2.28)$$

Each of these arrangements (macroscopically indistinguishable) is assumed to occur with the same *a priori* probabilities $p\{q_j\} : j = 1,\ldots,n$. They also bring out the physical significance of the prior probability assumption. They say that it is the 'equal-weight' postulate that allows the interpretation of q_j as the degeneracy for cell intensity j. Based on the uniform priors, q_j is considered proportional to the probability that n_j photons are distributed over z degrees of freedom. Assuming cell independencies, (i.e., the number of ways n_j photons coming from j^{th} cell are not influenced by the number of ways n_i ($i \neq j$) photons that come from i th cell), the total number of ways of forming the entire object is:

$$W(n_1, n_2, \ldots n_n) = \prod_{j=1}^{n} q_j(n_j) \qquad (2.29)$$

They write the probability of the object pattern as $p_j = n_j/n$, where $n = \sum_j n_j$. Applying Stirling's approximation to (2.28) and (2.29), they analyse the choice of different entropy forms based on the ratio n/z. Their conclusion is as follows: Bright radio objects with n/z >>1 , such as radio and X-ray astronomical objects should be processed with *logp* form of entropy. Less-bright objects with n/z<< 1 such as optical radioastronomical image restoration should be processed with the *plogp* form of entropy. Although their conclusions are considered significant, their analysis approach is specific in the context of image restoration, photon distributions and 'equal weight' prior probabilities. Cambridge *ME* researchers observe that by using the duration of observation, they convert the unquantized spatial power pattern into a quantized photon pattern and fail to provide full objectivity [51], [52]. Following the results of Kikuchi and Soffer, Frieden attempts to provide a more general basis for various forms of entropy.

2.5. FRIEDEN'S STATISTICAL APPROACH

R. Frieden, in another work, shows that the entropy based methods in general can be derived in a straightforward statistical approach [53], [54]. He sets out with the general definition of the probability of an event as $m_i/\boldsymbol{n}$, the ratio of the relative occurrences of the event to the total number of possible events, with the latter approaching infinity. In the present context, it is $p_i(= p(x_i))$ for the position of a photon and is given by (2.21). The objective of restoration is to find their values as applied to the data **d** in (2.2). The information about $\{p_i\}$ comes from two factors: First, $\sum_i p_i = 1$, and secondly, the prior knowledge about $\{p_i\}$, i.e., each event i has a probability q_i prior to knowing the data **d**. Let the prior probabilities for the events under consideration be $\{q_i\}$ with $\sum_i q_i = 1$. Frieden's approach is largely based on defining the prior knowledge using the 'multinomial probability law'. He establishes a general basis as follows: The photon emission process is an experiment repeated N times (N trials) with $\boldsymbol{n}$ possible outcomes. The quantity p_i stands for theprobability event of photon position for i^{th} pixel and is one of the $\boldsymbol{n}$ *distinct* events. The outcomes are assumed independent from trial to trial and from event to event. The probability of occurrence of any *one* of the N trials becomes $q_1^{m1} q_2^{m2} \ldots q_n^{mn}$. The number of ways the events can occur can be calculated using N and is given by

$$W_N(m_1, \ldots\ldots, m_n) \equiv \frac{N!}{m_1! \ldots\ldots m_n!} \tag{2.30}$$

Equation (2.30) may be compared with (2.7) to get a clear picture of Jayne's luminance model. Instead of taking the logarithm of (2.30) and immediately using Stirling's approximation as Jaynes did, Frieden makes use of (2.30) to define a *joint probability distribution*. As a result of prior probabilities $\{q_i\}$, the probability of any one set of ocurrences $\{m_i\}$ changes to

$$p_N(m_1,.....,m_n|\ q_1,.....,q_n) = \frac{n!}{m_1!.....m_n!}\ q_1^{m_1}.....\ q_n^{m_n} \qquad (2.31)$$

The product form of q_i's in (2.31) is the result of independence from trial to trial, and independence among the events. Let us assume that the prior probabilities follow a 'probability law' $p(q_1,.....,q_n)$. Also, assuming that $\{q_i\}$ are all present, (2.31) can be written using the partition law [$(p(x) = \int p(x|y)p(y)dy$] as

$$p(m_1,....,m_n) = \frac{n!}{m_1!...m_n!} \int ... \int \{q_1^{m_1}...q_n^{m_n}\}\ p(q_1,....,q_n)dq_1.....dq_n \qquad (2.32)$$

Maximizing (2.32) gives rise to the 'most likely' set of $\{p_i\}$ and Frieden calls (2.32) a 'general estimation principle'. Frieden also discusses in detail his viewpoint of prior knowledge. True prior knowledge according to Frieden is that which biases the estimate towards the true values $\{p_i\}$, and depends on the particular problem being solved. Since $\{p_i\}$ are the unknowns, it is difficult to assess the true $\{q_i\}$. However, the choice is expected atleast not to bias the estimate towards *incorrect* values. With this understanding he introduces a general form of prior probability law as

$$p(q_1,.....q_n) = \delta(q_1 - p_1)....\delta(q_n - p_n) \qquad (2.33)$$

and analyses the degree of 'maximum prior ignorance' in different entropy forms. For the uniform priors, $q_i = 1/n$ for $i = 1,.....,n$, the general estimation principle (2.32) becomes

$$\frac{N!}{m_1!....m_n!}\left(\frac{1}{n}\right)^{m_1 + m_2 + ...+ m_n} = \text{maximum.} \qquad (2.34)$$

But since $\sum_i m_i = N$ is a constant, the solution $\{m_i\}$ simply obeys

$$\frac{1}{m_1!....m_n!} = \text{maximum.} \qquad (2.35)$$

Now taking the logarithm of both sides and using Stirling's approximation, the principle

becomes

$$-\sum_{i=1}^{n} m_i \log m_i = \text{maximum} \qquad (2.36)$$

Substitiuting for $m_i = Np_i$ (2.36) becomes

$$-\sum_{i=1}^{n} p_i \log p_i = \text{maximum} \qquad (2.36.1)$$

which is Jayne's maximum entropy principle. Frieden's arguments follow the viewpoint that the form of entropy depends upon the prior knowledge conditions, which in turn depend upon the problem being solved. For photon occurrences from resolution cells, the degeneracy factors obey the Bose-Einstein statistics and the estimation principle is derived using (2.32) as

$$-\sum_{i=1}^{n} m_i \log m_i + \sum_{i=1}^{n} (m_i+z-1)\ln(m_i+z-1) = \text{maximum} \qquad (2.37)$$

where z is the number of quantum degrees of freedom for a photon in a cell given by (2.27) and $m_i = Np_i$. The principle (2.36.1) is well accepted on the grounds that the most of the *physical* phenomena obey the 'equal weights' postulate. With respect to the use of the form *plogp*, Frieden's analysis based on the prior knowledge estimation principles led to the same conclusion as that of Kikuchi and Soffer: when the object is of low intensity obeying $m_i << z$, the uniform prior probability principle applies. Using (2.32), he also provides a detailed discussion in [54] of the different entropy forms and the associated prior knowledge.

2.6. THE *ME* AS A CONSISTENT VARIATIONAL PRINCIPLE

Besides the arguments of choosing the appropriate form of entropy for image reconstruction, Kikuchi and Soffer assert that the form *logp* should be called entropy and

the other forms including Shannon's measure are only special cases of the *logp* form. Their ideas have raised objections in the literature, particularly in the context of finding **f** from (2.2). As a solution to this confusion, Shore and Johnson take a different information theoretic approach without invoking the concepts of Shannon's information measure [55], [56]. They start with the problem of inverting an equation

$$\int \mathbf{s}_r(\mathbf{x})\mathbf{p}^*(\mathbf{x})\,d\mathbf{x} = \bar{s}_r \qquad (2.38)$$

for known $\mathbf{s}_r(\mathbf{x})$ and s_r, $r = 1,\ldots,\mathbf{m}$. The problem is similar to (2.5), but stated as a continuous case. From specific information about $\mathbf{p}^*(\mathbf{x})$, the problem is to *infer* the function $\mathbf{p}^*(\mathbf{x})$ itself. Their approach is briefly outlined below following their own notations with minor modifications: Consider a system in a vector state $\mathbf{x} \in \mathbf{D}$ ($\mathbf{D}$ is a set of possible states). Let us suppose that $\mathbf{x}$ follows a probability density function $p(\mathbf{x})$. Let $\mathbb{D}$ (the font change may be noted) be the 'set' of all probability densities $\mathbf{p}$ on $\mathbf{D}$ such that $p(\mathbf{x}) \geq 0$ for $\mathbf{x} \in \mathbf{D}$ and

$$\int_{\mathbf{D}} \mathbf{p}(\mathbf{x})d\mathbf{x} = 1. \qquad (2.39)$$

The objective of inversion of (2.38) is to find a true, unknown state probability density $\mathbf{p}^* \in \mathbb{D}$. Let the information on $\mathbf{p}^*$ be in the form of linear equality constraints:

$$\int_{\mathbf{D}} \mathbf{p}^*(\mathbf{x})\,\mathbf{a}_k(\mathbf{x})d\mathbf{x} = \bar{a}_k. \qquad (2.40)$$

The constraint is an information denoted as $I = (\mathbf{p}^* \in \mathbb{1})$, where $\mathbb{1}$ is a constraint 'set', and $\mathbb{1} \subseteq \mathbb{D}$, which means that $\mathbf{p}^*$ is a member of the closed convex set $\mathbb{1} \subseteq \mathbb{D}$. Let $\mathbf{q} \in \mathbb{D}$ be some prior density that is an estimate of $\mathbf{p}^*$, prior to obtaining I such that $q(\mathbf{x} \in \mathbf{D}) > 0$. Given the prior q and the information I, the aim is to obtain a posterior density p by

minimizing a suitable functional H(p,q) over the constraint set $\mathfrak{I}$. That is

$$H(p,q) = \min_{p' \in \mathfrak{I}} H(p',q)$$

The question is: What should be H?. To make the discussions clear, they introduce a convenient notation p = q o I, meaning that the operator o takes a prior and new information to yield a posterior. To find the functional H, they introduce a set of consistency axioms *for the operator* o. The axioms are:

Uniqueness: The posterior p is unique for *any* prior $q \in \mathbf{D}$ and any *new* information $I = (p^* \in \mathfrak{I})$. The axiom implies that whatever may be the constraint and prior information the given *testable* information should remain the same.

Invariance: If $\mathfrak{J}$ is a coordinate transformation operator that changes $x \in \mathbf{D}$ to $y \in \mathbf{D}'$, then for any prior $q \in \mathbf{D}$ and new information $I = (p^* \in \mathfrak{I})$, $(\mathfrak{J}\, p) \text{ o } (\mathfrak{J}\, I) = \mathfrak{J}(p \text{ o } I)$ holds true. It means that when the problem is solved in two different coordinate systems, the posteriors should be related by the coordinate transformation.

System independence: Suppose there are two systems **D**1, **D**2 with probability densities $p1^* \in \mathfrak{I}1$ and $p2^* \in \mathfrak{I}2$ and the prior probability densities $q1 \in \mathbf{D}1$ and $q2 \in \mathbf{D}2$. The respective information for the systems are: $I1 = (p1^* \in \mathfrak{I}1)$ and $I2 = (p2^* \in \mathfrak{I}2)$ where $\mathfrak{I}1 \subseteq \mathbf{D}1$ and $\mathfrak{I}2 \subseteq \mathbf{D}2$. Then

$$(q1q2) \text{ o } (I1 \wedge I2) = (q1 \text{ o } I1)(q2 \text{ o } I2) \qquad (2.42)$$

Subset independence: Suppose the states **D** decompose naturally into disjoint subsets S_i, $i = 1,\ldots\ldots n$ and new information is obtained about the conditional probability densities $I_i = (p^* \mid S_i \in I_i)$ in each subset. Then

$$(q \circ I) \mid S_i = (q \mid S_i) \circ I_i. \tag{2.43}$$

The justifications for the axioms and mathematical proof for finding a suitable H can be found in [55]. The functional satisfying the axioms is found to be

$$H(p, q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{2.44}$$

which is called cross-entropy. They point out that the form (2.44) is due to Kullback and has many different names such as discrimination information, directed divergence and expected weight of evidence. The discrete equivalent of (2.44) is given by

$$S(p,q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right). \tag{2.44.1}$$

When q is a uniform distribution on a *finite* state space, minimizing the cross-entropy (2.44) is equivalent to the maximization of entropy (2.3). Their proofs also show that minimizing any function but entropy (refers to entropy as well as cross-entropy) will lead to inconsistency unless that function and entropy have identical maxima. The properties of (2.44) are reported in [56]. Thus Shore and Johnson have answered the question: Under what circumstances can a method of inference based on a variational principle be self-consistent? The resulting posterior has the form

$$p(x) = q(x) \exp\left\{ -\mu - \sum_{r=0}^{m} \lambda_r s_r(x) \right\}. \tag{2.45}$$

λ_r and μ are the Lagrange multipliers determined by (2.40) and (2.39) respectively. Their approach for the determination of prior probability distribution $q(x)$ can be found in [57]. Given the prior estimates q_j for the number of photons originating from the object point j, the prior distribution $q(x)$ must satisfy

$$q_j = \sum_{x} n_j p(x) \tag{2.46}$$

where n_j are the number of photons originating from point j. For example, q could be a

multivariate poisson distribution given by

$$q(\mathbf{x}) = \prod_{j=1}^{n} \exp\{-q_j\}\frac{(q_j)^{n_j}}{n_j!} \qquad (2.47)$$

Using (2.47) in (2.45) they derive the posterior object intensity as

$$\hat{f}_j = \sum_{n} n_j p(\mathbf{x}) = q_j \exp\left\{-\sum_{r=0}^{n} \lambda_r s_{rj}\right\} \qquad (2.48)$$

The Cambridge group researchers follow the view point of the *ME* method as a consistent variational technique along Shore and Johnson's line of approach [58], [52], [51], [23]. Their reasoning and interpretation however differ significantly. According to them, the *ME* method consists of choosing a single feasible image which has the greatest entropy given by (2.44), written as

$$S = -\sum_{i=1}^{n} p_i \log \frac{p_i}{m_i} \qquad (2.49)$$

Unlike Shore and Johnson, who attribute the form to Kullback, the Cambridge group refers to the form as Shannon/Jaynes entropy. Also, they interpret p_i as the probability corresponding to an *appropriate quality* of an image, and m_i as the *corresponding initial measure*. They understand the objective of finding **f** as the problem of restoring or reconstructing the 'shape' or the 'configurational structure' of an image. The objective is the same, whether the type of image is due to a 2-D light distribution, or 3 - D electrons in free space, or 1-D sequence of numbers such as absorptivity. An image is the intensity distribution $\{f_1,.....f_n\}$ of a physical object resulting from a statistical process. The intensity distribution represents the shape of an image. By removing its dimensionality by normalization, the intensity distribution becomes 'intensity proportions' $\{p_1,.....p_n\}$, which by themselves obey Kolmogorov's axioms of probability and hence can be considered as probability distributions. According to them, the question of 'where would

the next photon come from?' is a precise formal intepretation to identify the p_is as probabilistic quantities [58]. Stated simply, their arguments suggest that the physics of image generation is included in the intensity proportions and their entropy given by (2.49). They have also formulated the general consistency axioms of Shore and Johnson in terms of intensity proportions of an image without invoking the probabilistic arguments [22], [23].

As far as the image reconstruction problem is concerned, they say it does not matter whether the particles are photons or nuclei or quasars and also whether or not they obey classical or quantum statistics. They specifically object to the *logf* form of entropy *for* image reconstruction. As a justification, they derive the results of Burg, Kikuchi and Soffer directly by maximizing the Shannon/Jayne's entropy form using appropriate constraints and m_i [51], [52]. They also provide an analysis of their results in terms of classical and continuous objects, photons and electrons. Their experimental results bring out the consequences of using *logp* form of entropy for reconstruction with 1-D, 2-D and 3-D data. Their conclusions can be summarised as follows: Entropy does not measure the properties of a physical distribution but parametrises one's state of knowledge about a physical distribution. The choice of an entropy form for a particular application depends upon the 'question being asked'. For example, in time series analysis, it is the question of predicting an individual time sample or a spectral component. Burg's form, they say is appropriate for this purpose. On the other hand, if the problem is to reconstruct, that is to find the configurational information or the number of spectral lines, Shannon/Jayne's *plogp* entropy form is the appropriate one. In other words, Shannon/Jayne's solution is non-committal about the *position of* t from the distribution f(t). The Burg solution on the other hand, is non-committal about the time *samples* f(t) from an ensemble-average constrained distribution.

There exists still another viewpoint in the literature. It is argued that the *ME* method just fits the data with a non-linear transform of a band-limited function [59]. The functions

$F(B) = \ln B$ or $-B\ln B$ are just two members of a family of continuous entropy functions having the key properties $\partial^2 F/\partial B^2 < 0$ and $\partial^3 F/\partial B^3 > 0$. The best choice of F in a particular application, according to this view point depends on the user's requirements and not on fundamental issues. In another work, the *ME* method is interpreted as a regularizing functional for inverse problems [60], [20]. With the growing number of theoretical view points, interpretations and criticisms, the practical progress of the method becomes important to explore its prospects. Unfortunately, for images, the practical progress of the method, as we will see shortly is severely hampered by a more fundamental difficulty of obtaining a solution at all. In what follows the status of the method is summarised in the light of above discussions.

2.7. INVESTIGATION SUMMARY - I

Following the information theoretic approach explained under section 2.1 and 2.5, the *ME* method has a more general basis for inferential applications. On the basis of an *inference* view point, as Jaynes suggested, the consistency arguments alone are sufficient to determine the entropy formula given by (2.3) or (2.4). The point is further strengthened by Shore and Johnson in their axiomatic derivation of principles of entropy. Under the inference structure, the principle of entropy is simply the *maximization* or the *minimization* of the functionals, either (2.3) or (2.44.1) with the input information as constraints. The criterion for the choice of the functional depends upon the availability of a suitable prior probability distribution as applied to the problem being solved. The issue will be discussed further below. The extension of the inference principles to practical applications has been established by the Cambridge group researchers. They have applied the principles to a wide variety of application data, with promising results. They find all the results, analytical as well as experimental, irrespective of the type of application, including radio astronomical images center around the maximization of the functionals (2.44.1). Hence they introduce the term 'the consistent variational principle' for the applications based on the *ME*

principles.

The second school of thought opposes the general basis on the grounds of the physical arguments following the application of the entropy method to the radio astronomical image reconstruction by Wernecke and Addario. Observing the spectrum estimation basis for the reconstruction problem, they claim the use of *logf* form of entropy for images. Their work was followed by Kikuchi and Soffer to establish the physical basis and the general basis of the *logf* form for the entropy methods. Their arguments are defended by the information theorists on the grounds of the lack of an objective approach to support the general claims. R. Frieden follows a more general approach based on statistical principles to emphasize the importance of the form (2.44.1) and the role of the physical principles in determining the prior knowledge. In its present status, the functional form (2.44.1) is widely followed for images with uniform prior probabilities, since it has a well established basis for physical quantities [43], [45], [53], [61]. With uniform priors the functional form (2.44.1) reduces to (2.3).

While use of the form (2.3) continues in practice, research efforts are in progress to eliminate the arbitrariness in the usage of the prior probabilities q_i. At present, there is no well established way of assigning the probabilities to q, in relation with the quality or closeness of the final estimate to the true image. However, it is to the advantage of the method that its consistency basis allows this arbitrariness of prior knowledge to explore the various possibilities. Different possible forms of q and their meaning are discussed extensively by R. Frieden [54]. A recent survey shows a set of axioms proposed to generate q [22]. The connections of the *ME* method with the Bayesian approach also throws some light in this direction. Techniques for the formalization of relevant prior information (in the present context q) are being reported in many current applications that involve Bayesian problem formulation [27]. Following these advancements, the connection of the *ME* method with the Bayesian approach is receiving much attention in the current literature [62], [63]. The prior knowledge issue which is considered as a research issue in

the *ME* literature, is also subject to criticisms and confusions. This is partly because q is addressed by many different names depending on the context in which it appears. It includes terms such as initial model, prior probability distribution, prior information, hypothesis space and invariant measure space and so on. It appears that the form (2.3) with a uniform q is relatively free from these confusions for images.

This thesis adopts the view point that like the regularization theory for inverse problems, the *ME* theory in its present status requires progress in practical application areas. Within the framework of the mathematical properties of solutions to ill-posed and ill-conditioned problems, the progress of the regularization theory depends upon the information gained from the application studies as well [33]. Along the same line of reasoning, the physical and the statistical basis of the *ME* method need investigations in the light of applications. To be able to successfully use the method in applications, computational procedures and algorithms are necessary, an area which is not well established for the *ME* method in image processing. The computational status of the *ME* method is explored in the following.

2.8. COMPUTATIONAL SCHEMES

The computational phase of the *ME* method has a unique advantage: Irrespective of the varying conceptual basis, the method of finding the estimate of **f** remains the same. The method involves maximizing the entropy subject to the input constraints, which is a non-linear constrained optimization problem. A solution (*ME* estimate) is obtained using numerical procedures. Only very few procedures have been reported so far in the context of specific image processing applications. These known procedures are discussed in this section to explore the various computational difficulties.

2.8.1. Newton-Raphson technique

This is the earliest approach followed by Frieden to obtain the *ME* solution for image restoration. The technique involves solving a set of $m+1$ non-linear equations [11] using the Newton-Raphson procedure. He formulates the problem by making use of the weighted sum entropy (*flogf* form) written as

$$-\sum_{i=1}^{n} f_i \log f_i - \rho \sum_{i=1}^{n} e_i \log e_i \qquad (2.50)$$

where $e_j' = e_j + B$ (B is a constant employed to force the condition $e_j' > 0$). Maximization is carried out subject to the equalities

$$d_j = \sum_{k=1}^{n} A_{jk} f_k + e_j - B \qquad j = 1, \ldots, m \; m \leq n \qquad (2.50.1)$$

and

$$\sum_{i=1}^{n} f_i = T. \qquad (2.50.2)$$

T is the constant total intensity of the true image. The resulting solution is reported to be

$$\hat{f}_k = \exp\left(-1 + \mu + \sum_{j=1}^{m} \lambda_j A_{jk}\right) \qquad (2.51)$$

and

$$\hat{e}_j = \exp\left(-1 + \mu + \frac{\lambda_j}{\rho}\right) \qquad (2.51.1)$$

(2.51) and (2.51.1) are expected to satisfy the constraints (2.50.1) and (2.50.2) with f_j replaced by $\hat{f}_j$. The $m+1$ non-linear equations are solved for the unknowns λ_j, $j = 1, \ldots m$ and μ using the Newton-Raphson procedure. Frieden reports a satisfactory performance of the procedure, provided the values of B are large. His results showed significantly improved results in the restoration of a galactic image blurred by atmospheric turbulence [19]. The computational complexity of solving $m+1$ non-linear equations and the use of

special requirements for noise elimination make Frieden's approach less attractive.

2.8.2. Single constraint *ME* method

All the subsequent efforts for computational developments were based on the Bayesian approach explained under section 2.3. In obtaining the *a posteriori* distribution, the luminance frequency model is replaced by Frieden's statistical communication theory model of image formation. Using this model, the objective function J resulting from this model is the same as (2.16) but the frequencies are replaced by the pixel intensities. J is written as

$$J(\mathbf{f}) = S(\mathbf{f}) - \lambda Q(\mathbf{d},\mathbf{f}) \tag{2.52}$$

where

$$Q(\mathbf{d},\mathbf{f}) = \frac{1}{2}\sum_{j=1}^{m} \frac{\left| d_j - \sum_{i=1}^{n} A_{ji} f_i \right|^2}{\sigma_j^2} \tag{2.52.1}$$

$S(\mathbf{f})$ is given by (2.25.2). It may also be noted that unlike (2.16), the problem is stated using a Lagrange multiplier, which necessitates a constraint be introduced on the function Q to indicate the consistency of the *ME* estimate with the measured data. The most commonly used constraint is:

$$Q(\mathbf{d},\mathbf{f}) \cong \frac{m}{2} \tag{2.52.2}$$

The constraint is based on the central limit theorem [64], [65] stated as

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} \frac{e_j^2}{\sigma_j^2} = 1 \tag{2.52.2.1}$$

The *ME* method under the Bayesian approach has certain unique advantages on practical grounds. First, as already noted, the method is complemented with the noise reduction capabilities. Second, the use of a *single* Lagrange multiplier and the single constraint has the significant computational advantage of overcoming the complexity of finding $m+1$

Lagrange mutipliers. Maximizing the entropy subject to the Q-constraint in (2.52.2) is referred to as the **single constraint *ME* method** in this work.

2.8.3. Optimization techniques

This single constraint *ME* method is used by Wernecke and Addario for Fourier inversion in an image reconstruction application in radio astronomy [12]. Here, the problem is stated as an unconstrained maximization of

$$\int\int \log \hat{f}(x,y)dxdy - \lambda \sum_{k=1}^{m} \frac{|D_k - \hat{D}_k|^2}{\sigma_k^2} \qquad (2.53)$$

where D_k are the observed noisy samples of a two-dimensional transform of an unknown function f(x,y). $\hat{D}_k$ are the Fourier transforms of the estimate obtained using the relation

$$\hat{D}_k = \iint \hat{f}(x,y)\exp\{-j2\pi(u_k x + v_k y)\}dxdy. \qquad (2.53.1)$$

The Lagrange multiplier λ is chosen to satisfy the constraint

$$\sum_{k=1}^{m} \frac{|D_k - \hat{D}_k|^2}{\sigma_k^2} = m. \qquad (2.53.2)$$

Their paper besides introducing the *logf* form for the *ME* method for reconstruction applications, gives an account of the underlying computational difficulties in using optimization techniques. In their solution approach, the task involves one-dimensional search in an n-dimensional space. The simplest techniques are the steepest ascent and the conjugate gradient techniques by suitably defining the search directions in an iterative manner [66]. On one hand, the search is expensive, since a greater number of function and gradient evaluations is required for a more accurate answer. Reducing the number of searches may bring down the cost but may require more iterations for convergence. A major practical problem is the situation when $\partial Q/\partial f_j$ becomes negative. The resulting pixel value may become unrealistic by taking negative values. Several refinements have been

considered in their work to overcome the various practical problems. The difficulties of using optimization techniques have also been reported by many other researchers under different application contexts and a summarised account can be found in [21].

2.8.4. " Integral equation " approach

This approach was proposed by Gull and Daniell [13], [17], in image enhancement and image reconstruction from incomplete and noisy Fourier transform samples in radio and x-ray astronomy and enhancement. The problem involves maximizing the function

$$-\sum_{j=1}^{n} f_j \log f_j - \frac{\lambda}{2} \sum_{k \in A}^{m} \frac{|F_k - D_k|^2}{\sigma_k^2} \qquad (2.54)$$

where D_k and F_k are the Fourier transforms of the estimate and the observed data. respectively. (2.54) is similar to (2.53.1) except for the *flogf* form of entropy. Also they define the constraint in terms of χ^2 statistic . Instead of using optimization techniques to find the *ME* estimate, they use an equation approach. They determine the equation as

$$\hat{f}_j = \exp\left\{-1 + \lambda \sum_{k \in A} \frac{D_k - F_k}{\sigma_k^2} \exp\left\{\frac{2\pi i j k}{N}\right\}\right\} \qquad (2.54.1)$$

The solution is obtained using a general iterative approach. This approach is very simple and straightforward to obtain the *ME* solution but has been reported to be impractical for implementation on the grounds of numerical instabilities and poor convergence due to the exponential operator [16], [21]. Because of the simplicity of the approach, many attempts were made [14], [15], [67], [68], [69] to overcome the difficulties, but a satisfactory solution could not be found.

2.8.5. Entropy metric technique

The difficulty of obtaining solutions with the straightforward approach introduced algorithms with special control techniques [18]. Instead of searching along one direction at

a time, the search procedure is carried out in a subspace of three directions. The first direction vector maximises the entropy, while the second direction aims at reaching the desired value of Q. A control element is introduced through a metric tensor defining the lengths and angles in **f** space. The third vector is used as a correction vector. Based on these three direction vectors e1, e2 and e3, the iterative changes are made according to the relation

$$df = x_1 e_1 + x_2 e_2 + x_3 e_3. \qquad (2.55)$$

The coefficients x_1, x_2 and x_3 are determined using quadratic models for S and Q, based on the local gradients and curvatures. The technique has been reported to be successful and promising results have been reported in many restoration and reconstruction applications in crystallography.

2.8.6. Differential equation approach

In the differential equation approach, the solution is obtained by a 1-D search [70], [71]. The technique involves solving a set of differential equations using appropriate initial conditions, determined by the required constraints. Promising results have been obtained but no assessements have been made on the algorithm on the grounds of more work needed to optimize the search, time savings and improved termination criterion.

2.9. INVESTIGATION SUMMARY - II

The computational development of the *ME* method appears to be more severe than the fundamental progress of the method. The inference directly follows from the limited span of the computational techniques and the inverse problem applications, particularly machine vision applications. Even in its direct application area of image restoration and reconstruction, a simple and convincing approach has not yet evolved to support the usefulness of the method. The integral equation approach, the simplest technique of all, is

found unsuitable on the grounds of numerical instabilities and convergence problems. The computational status of the method for images therefore implies a more fundamental problem of obtaining a solution at all. Solution techniques with sophisticated control procedures are the only means at present to obtain *ME* solutions. The implementation details, the possible computational efficiency, the difficulties underlying these complex techniques are not known in clear terms on one hand, and on the other hand, their implications are too difficult to comprehend. Although superiority of the results is important, the computational aspects are more critical to the progress and acceptance of the method. Besides, the high dimensionality of images has a significant bearing on the computational efficiency of the *ME* algorithms in general. It is not uncommon at all to find the reported cpu time figures of the order of several minutes to hours for a *ME* algorithm.

There is a valid reason to account for the computational underdevelopment of the method. In general, non-linear techniques are theoretically well-praised for their merits, but are not put to use on the grounds of practical problems of computation. The *ME* method is not an exception to this shortcoming. Confronting this reality is important because, unlike many other methods, the *ME* method has no linear counterpart. Besides, in the computational areas, a non-linear constrained optimization problem in general has no general basis to overcome the associated optimality problems. Very recently simulated annealing techniques based on energy distribution concepts have been reported to be a satisfactory solution approach for complex optimization problems [72], [73]. In this sense, new solution techniques developed for the method in any of its application areas contribute to the increased understanding of non-linear solution methodologies in general. Also, the method being the same for all application areas, a solution approach developed for a particular area of application may be useful for other applications as well. Based on these arguments, it can be concluded that developing simple, economical and efficient working algorithms for the *ME* method in image processing are important for its progress.

The computational procedures appear to follow the standard formulations and well

established procedures in their applications. Experience however shows that the techniques do not always result in a stable and convergent solution. Investigations in general show two major pitfalls: lack of an analytical approach and the use of unjustified assumptions. The development of the solution procedures is not based on the understanding of *why* the convergence is poor or why the instability problems arise. As a consequence, the mechanism of a given procedure in obtaining the *ME* solution remains unknown. The understanding of a computational procedure in this direction is fundamental not only out of a practical necessity but also to effectively extend the procedure to different applications. The overlooked aspect is that the general computational procedures, in their extension to complex problems such as the constrained optimization problems need careful investigation, particularly with respect to their convergence aspects. Another overlooked aspect is the influence of assumptions used for computational convenience on the final solution. A typical example is the use of total constant intensity assumption without any accompanying justifications on its influence on the final solution. Indeed, none of the practical techniques explained above apply directly to the *plogp* form of the entropy. They are all based on the *flogf* form with the assumption of a known total constant intensity of the true signal, which is unknown in the problem. Also in using iterative procedures, the solution approach starts with an initial approximation and refined in successive steps towards a final solution. The error conditions of the procedure play an important role in determining the closeness of the successive estimates to the solution estimate. It is therefore important to analyse the given problem conditions and the applicability of control procedures with respect to the error conditions of the procedure. The problem of reducing the cpu execution time for the algorithm resides in parallellizing the algorithm suitable to a given parallel implementation scheme. Attempts have not been made in this direction for the *ME* algorithm so far. However, in a closely related approach based on Bayesian image models, parallel concepts are becoming increasingly popular using simulated annealing algorithms [25]. Following these observations, an attempt is made in the next chapter to

develop a general procedure starting from first principles, eliminating the assumptions wherever possible.

CHAPTER 3

DEVELOPMENT OF A PARAMETER CONTROLLED RELAXATION SCHEME

Preliminary studies of the computational procedures for the *ME* method have shown three major shortcomings to consider the development of a general procedure from first principles in this chapter. First, there are no established general schemes to obtain a *ME* solution. Even for specific applications related to data of type (2.2), experience has shown the computational procedures being either too complex to implement, or not feasible at all. Secondly, the procedures, for no clear reasons continue to make use of *flogf* form of entropy, when its usage, as discussed in chapter 2, is questionable on practical as well as fundamental grounds [12]. On practical grounds, the problem of maximizing the *plogp* form of entropy is not equivalent to the problem of maximizing the *flogf* form of entropy unless the total intensity of the true signal is known exactly, which is in fact an unknown in the problem. Thirdly, the constrained maximization problem of the *ME* method is often formulated as an unconstrained maximization problem with Lagrange multipliers in order to make use of general numerical solution procedures to obtain a *ME* solution. These schemes lack justifications or evidences of the feasibility to obtain a final solution and, more importantly, of the closeness of the final solution to the true *ME* solution.

This chapter proceeds with the derivation of a solution procedure starting from a constrained maximization formulation of the problem with *plogp* form of entropy without any assumptions on the total constant intensity. The choice of the *ME* method (single constraint or $\mathbf{m}$ - constraints), an application problem and the numerical procedure is not considered a major issue because of the following intuitive anticipations: Irrespective of the type of the *ME* method, the application and the numerical approach, the computational

problem of obtaining a *ME* solution remains the same as that of a non-linear constrained optimization problem. Therefore, the derivation of a solution procedure is bound to show up optimality problems for solution convergence in some form. The form and the techniques to be proposed to overcome the optimality problems will depend on the numerical procedure chosen for the study. The numerical procedure may also influence the functional aspects of the application class chosen. Based on the form and the techniques used, it should be possible to understand the general mechanism of convergence underlying the solution scheme. Infact all these issues will be encountered in the process of deriving a solution procedure for the *ME* method in the sections that follow.

3.1. PHASE - I: CONSTRAINED ITERATIVE PROCEDURE

In what follows, the single constraint *ME* method explained under section 2. 8.2 is stated as a constrained maximization problem and a solution is obtained in an equation form. The equation is also shown to be derivable using the Bayes' theorem explained under section 2.3.1. The equation is interpreted as a constrained iterative procedure in terms of initial, intermediate and termination conditions. With a suitable formulation of these conditions, the procedure is analysed for its influence on the application class.

3.1.1. Specifications for computations

For the computational problem under consideration, the input data are from an application class defined by (2.2), with the noise terms assumed to form a Gaussian distribution $\mathbf{N}(0,\sigma)$. For reasons which will be covered in subsequent sections, it is not necessary to know the standard deviation parameter σ of the noise distribution. Rewriting (2.2) we obtain

$$d_j = \sum_{k=1}^{n} A_{jk} f_k + e_j \quad j = 1, \ldots, m;\ m \le n. \tag{3.1}$$

Following the discussions in chapter 2, the true image vector **f** are associated with the shape distributions given by (2.21). The entropy of this distribution is written as

$$S(\mathbf{f}) = -\sum_{i=1}^{n} \frac{f_i}{\sum_{k=1}^{n} f_k} \log \frac{f_i}{\sum_{k=1}^{n} f_k}. \tag{3.2}$$

In writing (3.2), it is not assumed that Σf_k is a known constant. Alternative expression for (3.2) would be

$$S(\mathbf{f}) = -\left(\sum_{i=1}^{n} \frac{f_i}{\sum_{k=1}^{n} f_k} \log f_i \right) + \log \sum_{k=1}^{n} f_k \tag{3.2.1}$$

More concisely (3.2.1) can be expressed as

$$S(\mathbf{f}) = \log \left(\frac{\sum_{k=1}^{n} f_k}{q(\mathbf{f})} \right) \tag{3.2.2}$$

where

$$q(\mathbf{f}) = \exp \left(\sum_{i=1}^{n} \frac{f_i}{\sum_{k=1}^{n} f_k} \log f_i \right). \tag{3.2.3}$$

The function $Q(\mathbf{d},\mathbf{f})$ is defined as

$$Q(\mathbf{d},\mathbf{f}) = \frac{1}{2m} \sum_{j=1}^{m} \left(d_j - \sum_{k=1}^{n} A_{jk} f_k \right)^2. \tag{3.3}$$

Using (3.1), $Q(\mathbf{d},\mathbf{f})$ can also be written as

$$Q(\mathbf{d},\mathbf{f}) = \frac{1}{2m} \sum_{j=1}^{m} e_j^{\,2} = \frac{\sigma^2}{2}. \tag{3.3.1}$$

The computational problem as applied to (3.1) is now stated as follows: Given **d** and A, find the *ME* estimate of $\hat{\mathbf{f}}$ by maximizing the function

$$J\left(\mathbf{f}; \lambda_x, \sigma_x^2\right) = S(\mathbf{f}) - \lambda_x \left(\frac{2Q(\mathbf{d},\mathbf{f})}{\sigma_x^2} - 1 \right). \qquad (3.4)$$

The subscript x denotes an unknown quantity. Since the solution for (3.4) invariably involves iterations, the noise variance σ^2 may not be present in its true form. Hence, σ^2 has been replaced by an unknown σ_x^2 in writing (3.4). It is however assumed that σ_x^2 also results from a Gaussian distribution. The section 3.1.6 on the error analysis indeed justifies the Gaussian assumption. λ_x is the Lagrange multiplier introduced to satisfy the constraint

$$\frac{2Q(\mathbf{d}, \mathbf{f})}{\sigma_x^2} = 1 . \qquad (3.4.1)$$

The conditions for a single, unique and finite maximum of J in (3.4) are the same as those reported for J in (52) and (54) [69], [46] and are stated in Appendix-C. The removal of the total intensity assumption in the present approach introduces an additional condition $\log q(\mathbf{f}) \leq n/2$, apart from the standard $\lambda_x \geq 0$ condition.

3.1.2. Solution equation

The maximum point of $J(\mathbf{f})$ obeys the stationary point equation $\partial J(\mathbf{f}) / \partial \mathbf{f} = 0$ for the given solution conditions. From (3.4), the equation can be written as

$$\frac{\partial J\left(\mathbf{f}; \lambda_x, \sigma_x^2\right)}{\partial f_j} = \frac{\partial S(\mathbf{f})}{\partial f_j} - \frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d}, \mathbf{f})}{\partial f_j} = 0. \qquad (3.5)$$

Working out the derivatives (without assuming that the total intensity is a constant), it can be shown that

$$- \log f_j + \sum_{i=1}^{n} p_i \log f_i - \frac{2\lambda_x \sum_{k=1}^{n} f_k}{\sigma_x^2} \frac{\partial Q(\mathbf{d}, \mathbf{f})}{\partial f_j} = 0. \qquad (3.6)$$

Rearranging the terms yields

$$f_j = \exp\left\{\sum_{i=1}^{n} p_i \log f_i\right\} \exp\left\{ - \frac{2\lambda_x \sum_{k=1}^{n} f_k}{\sigma_x^2} \frac{\partial Q(\mathbf{d}, \mathbf{f})}{\partial f_j} \right\}. \qquad (3.7)$$

Using (3.2.3), (3.7) is expressed as

$$f_j = q(\mathbf{f}) \exp\left\{ - \frac{2\lambda_x \left(\sum_{k=1}^{n} f_k \right)}{\sigma_x^2} \frac{\partial Q(\mathbf{d}, \mathbf{f})}{\partial f_j} \right\} = q(\mathbf{f}) f_{j1} \qquad (3.8)$$

where

$$f_{j1} = \exp\left\{ - \frac{2\lambda_x \left(\sum_{k=1}^{n} f_k \right)}{\sigma_x^2} \frac{\partial Q(\mathbf{d}, \mathbf{f})}{\partial f_j} \right\}. \qquad (3.8.1)$$

Appendix-D shows how (3.8) can be derived using the Bayes' theorem. The pixel values of the solution estimate in (3.8) are all invariably positive due to the exponential operator, which is an important criterion for a true image estimate. In the discussions that follow, (3.8) is referred to as the **solution equation** and f_{j1} as the **kernel term.**

3.1.3. Interpretation

The solution equation is non-linear of form $\mathbf{x}=g(\mathbf{x})$ where $g(\mathbf{x}) = \exp(r(\mathbf{x}))$ with $\mathbf{x} \equiv \mathbf{f}$. The standard procedure involves starting from an initial estimate $\mathbf{x}^{(0)}$, at any iteration n, $g(\mathbf{x})$ is evaluated using the previous estimate $\mathbf{x}^{(n-1)}$. If $g(\mathbf{x})$ *exists*, it becomes an estimate for the iteration (n+1). Upon repeating the steps, the solution is reached at some iteration, say s, *provided the iterations converge*. This basic procedure evolves into a more

meaningful conceptual interpretation when extended suitably to the *ME* computational problem.

The process of carrying out the iterative procedure is visualized in terms of a 1-D path defined by the estimates calculated at various iterations. The path evolves in an $\mathbf{n}$-dimensional space in a direction determined by the successive estimates. The iterative function being exponential, is analytic everywhere, indicating that there exists an infinite number of paths to reach the required maximum point of $J(\mathbf{f}; \lambda_x, \sigma_x^2)$. Let $\hat{\mathbf{f}}^{(n)} = \{\hat{f}_1^{(n)}, \ldots\ldots, \hat{f}_{\mathbf{n}}^{(n)}\}$ be an estimate at any iteration n, called as an **intermediate estimate** at an **intermediate point** n of the path (n and $\mathbf{n}$ differ only by their font and style but stand for two different quantities: n refers to the number of iterations, while $\mathbf{n}$ represents the total number of pixel elements). The path traced by the iterative procedure up to the point n+1 is expressed in terms of the initial and intermediate estimates as

$$\mathbf{p}_{n+1} \rightarrow \hat{f}_j^{(0)}, \left\{\hat{f}_j^{(i)}\right\} \quad \text{for } i = 1, \ldots\ldots, n+1 \text{ and } j = 1, \ldots\ldots, \mathbf{n}. \qquad (3.9)$$

$\rightarrow$ denotes an assignment symbol. The superscript terms denote the intermediate points of the path. In obtaining the *ME* solution, the path is constrained by (3.4.1) with a suitable value for σ_x^2, or more appropriately by $\sigma_x^{2(i)}$. The constraint (3.4.1) with σ^2 replaced by $\sigma_x^{2(i)}$ is called an **intermediate constraint**. With the constraint incorporated, the iterative procedure becomes a **constrained iterative procedure**. The behaviour of the procedure in reaching the solution point is along this directed path is determined by λ_x and $q(\mathbf{f})$. Intuitively, the interpretation shows a necessity to control λ_x and σ_x^2 over iterations for a stable and a reliable solution. These ideas are conveyed by writing the path in (3.9) as

$$\mathbf{p}_{n+1} \rightarrow \hat{\mathbf{f}}^{(0)}, \left\{F\left(q(\hat{\mathbf{f}})^{(i)}, \lambda_x^{(i)}, \sigma_x^{2(i)}, \frac{\partial Q\left(\mathbf{d}, \hat{\mathbf{f}}^{(i)}\right)}{\partial \mathbf{f}}\right)\right\} \quad \text{for } i = 1, \ldots\ldots, n+1. \qquad (3.10)$$

F is a function given by (3.8) with f_j replaced by $\hat{f}_j^{(i)}$ and also $q(\hat{\mathbf{f}})^{(i)}$ replacing $q(\mathbf{f})$. The **destination point** of the path is fixed by the solution estimate denoted as $\hat{\mathbf{f}} = \{\hat{f}_1,.....,\hat{f}_n\}$. It is characterized by specific $\lambda_x^{(i)}$ and $\sigma_x^{2(i)}$, satisfying certain **termination conditions**, in addition to the intermediate constraints. These conditions will be discussed at the end of this section. The following section gives the quantitative definition of various terms introduced in this section.

3.1.4. Initial condition

The initial condition of the procedure in the present study is considered as a maximally uncertain situation. This situation is marked by the beginning of the procedure, when there are no constraints, that is when $\lambda_x^{(0)} = 0$ in (3.8). The solution for this condition is obtained by the unconstrained maximiziation of the entropy resulting in a uniform estimate given by

$$\hat{f}_j^{(0)} = \exp\{-1\} \quad j = 1,\cdots\cdots n. \tag{3.11}$$

From the above approach of finding the initial estimate, it may be noted that the initial estimate comes from the prior knowledge term of the general *plogp/q* entropy form given by (2.44.1) in chapter 2. Since the prior knowledge is assumed uniform, as implied by the *plogp* form of entropy, the maximization has resulted in (3.11). With the initial condition fixed, the path (3.10) simplifies to

$$\mathbf{p}_{n+1} \rightarrow \{\hat{f}_j^{(i)}\} \quad \text{for } i = 0,\cdots\cdots,n+1 \text{ and } j = 1,\cdots\cdots,n. \tag{3.12}$$

3.1.5. Intermediate variables and conditions

The intermediate estimate for the pixel j at any intermediate point n+1 is obtained simply by replacing f_j and $q(\mathbf{f})$ in (3.8) by $\hat{f}_j^{(n)}$ and $q(\hat{\mathbf{f}})^{(n)}$ respectively. It is given by

$$\hat{f}_j^{(n+1)} = q(\hat{\mathbf{f}})^{(n)} \exp\left\{ -\frac{2\lambda_x^{(n)}\left(\sum_{k=1}^{n} \hat{f}_k^{(n)}\right)}{\sigma_x^{2(n)}} \frac{\partial Q^{(n)}}{\partial f_j} \right\} \tag{3.13}$$

where

$$q(\hat{\mathbf{f}})^{(n)} = \exp\left\{ \sum_{j=1}^{n} \hat{p}_j^{(n)} \log \hat{f}_j^{(n)} \right\} \tag{3.13.1}$$

with

$$\hat{p}_j^{(n)} = \frac{\hat{f}_j^{(n)}}{\sum_{k=1}^{n} \hat{f}_k^{(n)}} . \tag{3.13.2}$$

In (3.13), $Q(\mathbf{d},\hat{\mathbf{f}}^{(n)})$ is written as $Q^{(n)}$ for clarity. Since the vector $\mathbf{d}$ is invariant over n, it implies that $Q^{(n)} \equiv Q(\hat{\mathbf{f}}^{(n)}) \equiv (Q(\hat{\mathbf{f}}))^{(n)}$. $Q^{(n)}$ follows from (3.3) with $\mathbf{f}$ replaced by $\hat{\mathbf{f}}^{(n)}$ and is given by

$$Q^{(n)} = \frac{1}{2m} \sum_{j=1}^{m} \left[d_j - \sum_{k=1}^{n} A_{jk} \hat{f}_k^{(n)} \right]^2 = \frac{1}{2m} \sum_{j=1}^{m} \hat{\zeta}_j^{(n)2} . \tag{3.14}$$

The term $\zeta_j^{(n)}$ is the **intermediate estimate error** term for pixel j at iteration n given by

$$\hat{\zeta}_j^{(n)} = d_j - \sum_{k=1}^{n} A_{jk} \hat{f}_k^{(n)} . \tag{3.15}$$

$\partial Q^{(n)}/\partial f_j$ is a vector of n elements obtained from (3.14) as

$$\frac{\partial Q^{(n)}}{\partial f_j} = \frac{1}{m} \sum_{r=1}^{n} \left(d_r - \sum_{k=1}^{n} A_{rk} \hat{f}_k^{(n)} \right) A_{rj} . \tag{3.16}$$

The parameters $\lambda_x^{2(n)}$ and $\sigma_x^{2(n)}$ will have to be determined. The determination of $\sigma_x^{2(n)}$, as mentioned earlier, is based on the intermediate constraint (3.4.1) with Q replaced by $Q^{(n)}$ written as

$$\frac{2Q^{(n)}}{\sigma_x^{2(n)}} = 1. \qquad (3.17)$$

(3.17) differs from (3.4.1) by the numerator term. In (3.4.1), the value of Q is exact, whereas in (3.17) it is the value as determined by the procedure at iteration n based on the procedure conditions. An error analysis of the procedure is carried out in the following section to evaluate the form of σ_x^2 in the procedure.

3.1.6. Error analysis and noise variance elimination

Rewriting the intermediate estimate error given by (3.15)

$$\hat{\zeta}_j^{(n)} = d_j - \sum_{k=1}^{n} A_{jk} \hat{f}_k^{(n)}. \qquad (3.18)$$

Let the difference(error) between the n^{th} intermediate estimate and the true solution be denoted by a random variable $\varepsilon_k^{(n)}$. Equation (3.18) can then be written as

$$\hat{\zeta}_j^{(n)} = d_j - \sum_{k=1}^{n} A_{jk} \left(f_k + \hat{\varepsilon}_k^{(n)} \right). \qquad (3.19)$$

If the input data are subject to noise, (3.19) becomes

$$\hat{\zeta}_j^{(n)} = e_j - \sum_{k=1}^{n} A_{jk} \hat{\varepsilon}_k^{(n)} \qquad (3.20)$$

where the terms e_j are derived directly from (3.1). Equation (3.20) can be expressed as

$$\hat{\zeta}_j^{(n)} = e_j - \hat{\xi}_j^{(n)} \qquad (3.21)$$

where

$$\hat{\xi}_j^{(n)} = \sum_{k=1}^{n} \hat{\varepsilon}_k^{(n)} A_{jk}. \qquad (3.21.1)$$

(3.21) shows that the intermediate estimate error for pixel j is the difference between of the noise error terms in the input data and the estimation error terms in (3.21.1) due to the

iterative procedure. The estimation error terms result from a summation of a random variable $\varepsilon_k^{(n)}$ over n elements using A_{jk}. It follows from the central limit theorem [64], [65] that at any intermediate point n, when n is large, which is usually the case with images, irrespective of the type of the distribution of the individual random errors ε_k, the estimation error terms approximate a Gaussian distribution. Since the noise terms in the input data are additive and Gaussian in the problem under study, it follows that the intermediate error at any point n is also Gaussian. Similar arguments hold true in the absence of noise in the input data. This is because in the absence of noise, the intermediate error term becomes

$$\hat{\zeta}_j^{(n)} = \sum_{k=1}^{n} \hat{\varepsilon}_k^{(n)} A_{jk} \tag{3.22}$$

which shows that the intermediate error terms still form a Gaussian distribution through estimation error terms. In other words, the procedure has become 'transparent' to the input noise error terms through the intermediate estimate error terms as a result of the iterations. The **intermediate error variance** is written as

$$\hat{\sigma}_\zeta^{2(n)} = \frac{\sum_{j=1}^{m} \hat{\zeta}_j^{2(n)}}{m} \tag{3.23}$$

By squaring and summing (summation over the pixels) the terms of (3.21) on both sides, the variance (3.23) can be expressed as

$$\hat{\sigma}_\zeta^{2(n)} = \sigma^2 + \hat{\sigma}_\xi^{2(n)} - 2\hat{\sigma}_{g\xi}^{(n)} \tag{3.23.1}$$

where

$$\hat{\sigma}_{g\xi}^{(n)} = \frac{1}{m} \sum_{j=1}^{m} e_j \hat{\xi}_j^{(n)} \tag{3.23.2}$$

The possibility of reducing the noise variance using σ_x^2 is explored below.

3.1.7. Determination of σ_x^2

In the procedure, the right hand side term of (3.23) is calculated as $Q^{(n)}$ (equation 3.14). Hence, (3.23) becomes

$$\hat{\sigma}_\zeta^{2(n)} = 2Q^{(n)}. \tag{3.24}$$

Combining (3.24) and (3.17) it can be seen that the intermediate constraint is satisfied in the procedure if

$$\sigma_x^{2(n)} = \hat{\sigma}_\zeta^{2(n)}. \tag{3.24.1}$$

The situation indeed shows consistency with a reported work [38], where a result of form (3.24.2) appears as a consequence of applying Bayes' theorem using Jefferey's prior probability distribution for σ (a completely unknown parameter) in the *ME* method. Substitution of (3.21) in (3.16) gives rise to

$$\frac{\partial Q^{(n)}}{\partial f_j} = \frac{1}{m} \sum_{r=1}^{n} \hat{\zeta}_r^{(n)} A_{rj}. \tag{3.25}$$

Using (3.24.1) and (3.25) in the solution equation (3.13), we get

$$\dot{f}_j^{(n+1)} = q(\hat{f})^{(n)} \exp\left(- \frac{2\lambda_x^{(n)} \left(\sum_{k=1}^{n} \dot{f}_k^{(n)} \right)}{\hat{\sigma}_\zeta^{2(n)}} \sum_{r=1}^{n} \hat{\zeta}_r^{(n)} A_{rj} \right) \tag{3.26}$$

The intermediate error variance term in the denominator term of (3.26) is given by (3.23), which shows that the noise variance (if any) automatically gets eliminated in the solution. Besides, the intermediate constraint in (3.17) is always satisfied and the path is constrained as required by the computational problem.

3.1.8. Termination condition(s)

The condition(s) employed by the procedure to identify a specific intermediate point

close to the solution point, are called termination condition(s). The solution point is fixed by an estimate that satisfies (3.13), which happens when the iterations converge to their limiting condition. Under the assumption that the iterations converge, the convergence condition is the same as that of any general iterative procedure given by

$$\lim_{n\to\infty}\left(\hat{f}_j^{(n+1)} - \hat{f}_j^{(n)}\right) = 0. \tag{3.27}$$

The procedure requirements for convergence depend on further developments that are to follow. Assuming for the present that Q decreases with increasing iterations in the procedure, we can see from (3.23) and (3.24.1) and (3.17) that the intermediate constraint tends to its theoretical constraint (2.52.2), encountered earlier in chapter2. The convergence of the constraint can be expressed as

$$\lim_{n\to\infty}\frac{2Q^{(n)}}{\hat{\sigma}_\varepsilon^{2(n)}} = \frac{2Q}{\sigma^2} = 1\,. \tag{3.28}$$

Because of the finiteness of the intermediate error terms, the denominator is a non-zero quantity for noisy as well as noiseless cases. It can be seen from (3.23) and (3.14) that the absolute value of Q near the solution point is influenced by the level of noise. For the no-noise case, the theoretical limit can be written as

$$\lim_{n\to\infty} Q^{(n)} = 0. \tag{3.28.1}$$

In presence of noise, the limit is expected to be

$$\lim_{n\to\infty} Q^{(n)} = \frac{\sigma^2}{2}. \tag{3.28.2}$$

The intermediate variance reduces by a factor of m, when it is true that

$$\lim_{n\to\infty} Q^{(n)} = \frac{m}{2} \tag{3.28.3}$$

As we will see later, the limiting conditions (3.28) can be satisfied in the procedure using

suitable techniques. It follows therefore that the convergence condition (3.27) is the only termination condition necessary and sufficient for the procedure.

3.1.9. Determination of λ_x and its difficulties

With $\sigma_x^{2(i)}$ given by (3.24.2), the path becomes solely a function of λ_x written as

$$p_{n+1} \rightarrow \left\{\hat{f}_j^{(i)}\right\} = \left\{F(\lambda_x)\right\} \quad \text{for } i = 0,\cdots\cdots,n+1 \text{ and } j = 1,\cdots\cdots,n \tag{3.29}$$

where the estimate pixels $\hat{f}_j^{(i)}$ are obtained using

$$\hat{f}_j^{(i)} = \hat{q}^{(i)} \exp\left\{-\frac{2\lambda_x^{(i)}}{\hat{\sigma}_\zeta^{2(i)}} \frac{\partial Q^{(i)}}{\partial f_j}\right\} \tag{3.30}$$

(3.30) is written without the total intensity term in the numerator as a general form refering to the two types of solution approaches: With and without involving the assumption of the total intensity being a constant. The derivation in appendix-D refers to the solution equation with the assumption. This thesis work, which is the case of not involving the assumption refers to the solution equation given by (3.13). The equivalence of (3.30) to (3.13) will be addressed shortly in the context of a relaxation scheme under section 3.3. In writing (3.30), $q(\hat{\mathbf{f}})^{(i)}$ is replaced by $\hat{q}^{(i)}$ for the purpose of clarity. Also, λ_x can be combined with the intermediate variance and written as

$$\lambda^{(i)} = \frac{2\lambda_x^{(i)}}{\hat{\sigma}_\zeta^{2(i)}} = \frac{\lambda_x^{(i)}}{Q^{(i)}} \tag{3.31}$$

The main purpose of λ (the suffix x has been dropped for convenience) in the procedure is to satisfy the limits on Q in (3.28.1) or (3.28.2). Also, by reaching these limits faster, the computer processing time can be considerably reduced. This requires the values of λ to be as high as possible, so that Q is lowered as fast as possible. With increasing values of λ,

the feasibility of the procedure however becomes questionable, because of the exponential operator in (3.30). It leads to large values of the estimates, leading to integer overflow and underflow. The exponential instability in turn will lead to serious iterative convergence and termination difficulties. This behaviour is best demonstrated in terms of the characteristics of the iterative function.

For an iterative equation of form $\mathbf{x}=g(\mathbf{x})$, the sufficient condition for convergence is

$$|g'(\mathbf{x})| < 1 \tag{3.32}$$

where $g'(\mathbf{x})$ is the first derivative of the iterative function $g(\mathbf{x})$. If (3.32) is satisfied, then the iteration must converge for all x, including all the intermediate values of x_j and the true value x [74], [75]. The iterative function at point n for pixel j follows from (3.30) and is written as

$$g(\hat{f}_j^{(n)}) = \hat{q}^{(n)} \exp\left\{ -\lambda^{(n)} \frac{\partial Q^{(n)}}{\partial f_j} \right\}. \tag{3.33}$$

$\lambda^{(n)}$ is given by (3.31) with i replaced by n. The magnitude of the derivative of (3.33) follows the relationship

$$\left| g'(\hat{f}_j^{(n)}) \right| \propto \lambda^{(n)} \hat{f}_j^{(n)} . \tag{3.34}$$

With increasing values of λ and $\hat{f}_j^{(n)}$, we can see from (3.34) that the convergence condition in (3.32) is more likely to be violated. To characterize the estimate $\hat{f}_j$ with this behaviour, the symbol - is appended over the estimate symbol ^ and is named as the **unbounded intermediate estimate.** The improvement lies in choosing λ as high as possible to reach the limit on Q faster, but at the same time, not high enough to cause exponential instability. Obviously, these conflicting requirements are very difficult to meet using empirical techniques. A quantitative approach is necessary to improve the situation.

3.2. PHASE II: THE CONSTRAINED ITERATIVE PROCEDURE IN A RELAXATION SCHEME

One possible way of overcoming the conflicting situation with λ is to introduce additional steps for its iterative convergence. The standard way of improving convergence in an iterative procedure is to use a relaxation parameter β to average the successive estimates. This step by itself can not result in the convergence improvement. As will be seen below, this step, apart from smoothing the intermediate estimates, exercises its influence on the intermediate variables of the procedure at different levels.

3.2.1. Relaxation of intermediate estimates using β

By using the standard procedure to improve convergence, the successive estimates (3.30) are averaged using a parameter $\beta^{(n)}$. As a result, the intermediate estimate corresponding to point n+1 becomes

$$\bar{f}_j^{(n+1)} = \left(1-\beta^{(n+1)}\right)\bar{f}_j^{(n)} + \beta^{(n+1)}\bar{f}_j^{(n+1)} \qquad (3.35)$$

with

$$0 < \beta^{(i)} \leq 1, \qquad \text{for } i = 1, \cdots\cdots, n \qquad (3.35.1)$$

and

$$\bar{f}_j^{(n+1)} = \hat{q}^{(n)} \exp\left\{ -\lambda^{(n)} \frac{\partial Q^{(n)}}{\partial \bar{f}_j} \right\} \qquad (3.35.2)$$

$\hat{q}^{(n)}$ is now calculated in terms of the estimates given by (3.35) as

$$\hat{q}^{(n)} = \exp\left\{ \sum_{j=1}^{n} \bar{p}_j^{(n)} \log \bar{f}_j^{(n)} \right\} \qquad (3.35.3)$$

with

$$\bar{p}_j^{(n)} = \frac{\bar{f}_j^{(n)}}{\sum_{k=1}^{n} \bar{f}_k^{(n)}} \qquad (3.35.4)$$

In (3.35.2), $Q^{(n)} = Q(\bar{\bar{f}}^{(n)})$ and $\bar{\bar{f}}^{(n)}$ can be called as the **relaxed intermediate estimate**. Equation (3.35) is a recursive relation which can be substituted successively for its n^{th} estimate in the right hand side resulting in

$$\bar{f}_j^{(n+1)} = \frac{1}{n+1} \sum_{i=0}^{n+1} (n+1)\beta_p^{(i,\,n+1)} \bar{f}_j^{(i)} \qquad (3.36)$$

where

$$\beta_p^{(i,\,n+1)} = \left(1-\beta^{(n+1)}\right)\left(1-\beta^{(n)}\right)\ldots\ldots\left(1-\beta^{(i+1)}\right)\beta^{(i)} \qquad (3.36.1)$$

In (3.36), for notational convenience, it is assumed that $\beta^{(0)} = 1$ and when $i = n+1$, $\beta_p^{(n+1,n+1)} = \beta^{(n+1)}$. Also, n+1 is introduced in the numerator and the denominator for analytical purpose. Expressing (3.36) in a simpler form

$$\bar{f}_j^{(n+1)} = \frac{1}{n+1} \sum_{i=0}^{n+1} \bar{f}_j^{(i,\,n)} \qquad (3.37)$$

where

$$\bar{f}_j^{(i,\,n)} = (n+1)\,\beta_p^{(i,\,n+1)} \bar{f}_j^{(i)} \qquad (3.37.1)$$

(3.37) implies that the estimates are scaled using $\beta_p^{(i,n+1)}$ as in (3.37.1) and averaged over iterations in the process of relaxation.

3.2.2. Influence of relaxation on the procedure

Substituting (3.37) in (3.14), and denoting the result as $\bar{\bar{Q}}$ yields

$$\bar{\bar{Q}}^{(n)} = \frac{1}{2m} \sum_{j=1}^{m} \left[\left(d_j - \frac{1}{n} \sum_{i=0}^{n} \sum_{k=1}^{n} A_{jk} \bar{f}_k^{(i,\,n)} \right) \right]^2 \qquad (3.38)$$

We can make use of n in the denominator to write (3.38) as

$$\overline{Q}^{(n)} = \frac{1}{2m}\sum_{j=1}^{m}\left[\frac{1}{n}\sum_{i=0}^{n}\left(d_j - \sum_{k=1}^{n} A_{jk}\,\bar{\bar{f}}_k^{(i,n)}\right)\right]^2 \qquad (3.39)$$

Identifying the term within the inner parenthesis as the new error term, (3.39) is simplified to

$$\overline{Q}^{(n)} = \frac{1}{2m}\sum_{j=1}^{m}\bar{\bar{\zeta}}_j^{(n)\,2} \equiv \left[\frac{1}{2m}\sum_{j=1}^{m}\bar{\bar{\zeta}}_j^{2}\right]^{(n)} \qquad (3.40)$$

where

$$\bar{\bar{\zeta}}_j^{(n)} = \frac{1}{n}\sum_{i=0}^{n}\bar{\bar{\zeta}}_j^{(i,n)} \qquad (3.41)$$

with

$$\bar{\bar{\zeta}}_j^{(i,n)} = d_j - \sum_{k=1}^{n} A_{jk}\,\bar{\bar{f}}_k^{(i,n)} \qquad (3.41.1)$$

Also, $\partial Q/\partial f_j$ becomes

$$\frac{\partial \overline{Q}^{(n)}}{\partial f_j} = \frac{1}{nm}\sum_{i=0}^{n}\left(\sum_{r=1}^{n}\left(d_r - \sum_{k=1}^{n} A_{rk}\,\bar{\bar{f}}_k^{(i,n)}\right)A_{rj}\right) \qquad (3.42)$$

(3.42) can be expressed in a much simpler form as

$$\frac{\partial \overline{Q}^{(n)}}{\partial f_j} = \frac{1}{n}\sum_{i=0}^{n}\frac{\partial Q^{(i,n)}}{\partial f_j} \qquad (3.43)$$

where

$$\frac{\partial Q^{(i,n)}}{\partial f_j} = \frac{1}{m}\sum_{r=1}^{n}\left(d_r - \sum_{k=1}^{n} A_{rk}\,\bar{\bar{f}}_k^{(i,n)}\right)A_{rj} \qquad (3.43.1)$$

The error analysis results obtained for the intermediate estimates in section 3.1.4 hold true for the relaxed intermediate estimates as well. The intermediate variance resulting from the relaxation scheme is estimated by

$$\bar{\bar{\sigma}}_\zeta^{2(n)} = \frac{\left[\sum_{j=1}^{m}\bar{\bar{\zeta}}_j^{2}\right]^{(n)}}{m} = 2\,\overline{Q}^{(n)} \qquad (3.44)$$

As a result of the change in Q, λ in (3.31) gets modified to

$$\lambda^{(n)} = \frac{\lambda_x^{(n)}}{\bar{Q}^{(n)}} \qquad (3.45)$$

The path becomes a function of the 'relaxation parameter' β. That is

$$p_{n+1} \rightarrow \left\{ \bar{f}_j^{(i)} \right\} = \left\{ F\left(\lambda^{(i)}, \beta^{(i)}\right) \right\} \text{ for } i = 0, \cdots\cdots, n+1 \text{ and } j = 1, \cdots\cdots, n. \qquad (3.46)$$

$\bar{\hat{f}}_j^{(i)}$ is obtained using (3.35) with the unbounded estimate given by

$$\bar{f}_j^{(i)} = \hat{q}^{(i)} \exp\left\{ -\lambda^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \right\} \qquad (3.46.1)$$

The termination condition (3.28) is obtained in terms of $f_j^{(n)}$. With this understanding, the convergence situation is explored in the following sections. Before proceeding with this step, the form of $\lambda^{(n)}$ can be modified to suit the practical requirements.

3.2.3. Practical form for λ_x

As noted earlier, λ is expected to be as high as possible, so that, the termination convergence of Q is faster with less computer processing time. Also, the λ variation should guarantee error reduction in the procedure. These conditions are satisfied simply by forcing λ to strictly increase with iterations. One way of increasing λ is to iterate it in the same way as the relaxed intermediate estimates, so that

$$\bar{\lambda}^{(n)} = \bar{\lambda}^{(n-1)} + \lambda^{(n)} = \sum_{i=0}^{n} \lambda^{(i)} \qquad (3.47)$$

The initial value, $\lambda^{(0)}$ is assumed to be 0. Substitituting for $\lambda^{(i)}$ from (3.45) and considering λ_x to be a positive constant K_λ

$$\bar{\lambda}^{(n)} = \sum_{i=0}^{n} \frac{K_\lambda}{\bar{Q}^{(i)}} \qquad (3.48)$$

Because of the summation of strictly positive quantities, $\bar{\lambda}$ will always be an increasing parameter with iterations so that $\bar{Q}$ always decreases [since $d\bar{Q}/d\lambda$ is negative, explanations follow in chapter 4]. This means that in the procedure, the intermediate error-variance is a decreasing quantity, making the limiting condition definite on $\bar{Q}$. Also, the convergence becomes faster compared with the λ variation given by (3.31), since $\bar{\lambda}^{(n)} > \lambda^{(n)}$. With $\bar{\lambda}^{(n)}$ given by (3.47), the path becomes

$$\mathbf{p}_{n+1} \rightarrow \begin{vmatrix} \bar{\hat{f}}_j^{(i)} \\ \beta^{(i)} \end{vmatrix} = \begin{vmatrix} \Gamma\left(\bar{\lambda}^{(i)}, \beta^{(i)}\right) \end{vmatrix} \quad \text{for } i = 0, \cdots\cdots, n+1 \text{ and } j = 1, \cdots\cdots, \mathbf{n}. \qquad (3.49)$$

$\bar{\hat{f}}_j^{(i)}$ follows from (3.36) as

$$\bar{f}_j^{(i)} = \sum_{w=0}^{i} \beta_p^{(w,i)} \bar{f}_j^{(w,i)} \qquad (3.50)$$

with

$$\bar{f}_j^{(w,i)} = \hat{q}^{(w)} \exp\left\{ -\bar{\lambda}^{(w)} \frac{\partial \bar{Q}^{(w,i)}}{\partial f_j} \right\} \qquad (3.50.1)$$

where

$$\bar{\lambda}^{(w)} = \sum_{r=0}^{w} \lambda^{(r)} \qquad (3.50.2)$$

and

$$\frac{\partial \bar{Q}^{(w,i)}}{\partial f_j} = \frac{1}{w} \sum_{s=0}^{w} \frac{\partial Q^{(s,w)}}{\partial f_j} \qquad (3.50.3)$$

The procedure now has all its variables defined except β and the term K_λ in $\bar{\lambda}$. Before proceeding to determine them,the convergence sensitivity of β is explored to study its significance in the procedure. An analysis in this direction is also expected to provide insight into the inherent difficulties of the approach.

3.2.4. Significance of using β

Let us first consider the simple case of using $\beta^{(n)}$ with λ variations given by (3.45). We know from our analysis that $\beta^{(n)}$ influences the procedure by averaging $\partial Q/\partial f_j$ as in (3.50.3). The consequence of this influence can be seen by substituting (3.43) in (3.46.1) and expressing the result as

$$\tilde{f}_j^{(n)} = \hat{q}^{(n)} \prod_{i=0}^{n} \exp\left\{ -\lambda^{(n)} \frac{\partial Q^{(i)}}{\partial f_j} \right\} = \hat{q}^{(n)} \prod_{i=0}^{n} \tilde{f}_{j1}^{(n,i)} \qquad (3.51)$$

where the kernel term is

$$\tilde{f}_{j1}^{(n,i)} = \exp\left\{ -\lambda^{(n)} \frac{\partial Q^{(i)}}{\partial f_j} \right\} \qquad (3.51.1)$$

In (3.51) the kernel terms, calculated with λ at iteration n and $\partial Q/\partial f_j$ at iteration w are multiplied. With increasing iterations, the number of kernel terms increases in the product, the magnitude of each term being dependent upon the magnitude of the exponential argument $\lambda^{(n)}\partial Q^{(i)}/\partial f_j$. The unbounded behaviour of the *ME* estimate is caused by the exponential growth of the argument with increasing values of λ. It leads to integer overflow or underflow in the procedure. The situation can be expected to improve by using β for scaling and averaging in the relation (3.36), rewritten as

$$\tilde{f}_j^{(n)} = \sum_{i=0}^{n} \beta_p^{(i,n+1)} \tilde{f}_j^{(i)} \qquad (3.51.2)$$

With increasing iterations, the number of terms under summation increases along with the number of product terms, the two being controlled by β. Although the scheme shows flexibility for convergence improvement, λ is not used in its practical form for computational speed improvements.

Next, by considering the case with λ for computational speed improvements, but without β, the significance of β can be demonstrated. This situation was encountered earlier

in section 3.1.9, where we emphasized the need for relaxation, but a different line of reasoning was followed. Substituting (3.50.2) in (3.33) yields

$$\hat{f}_j^{(n)} = \hat{q}^{(n)} \prod_{i=0}^{n} \exp\left\{ -\lambda^{(i)} \frac{\partial Q^{(n)}}{\partial f_j} \right\} = \hat{q}^{(n)} \prod_{i=0}^{n} \hat{f}_{j1}^{z(i,n)} \qquad (3.52)$$

with the kernel term

$$\hat{f}_{j1}^{z(i,n)} = \exp\left\{ -\lambda^{(i)} \frac{\partial Q^{(n)}}{\partial f_j} \right\} \qquad (3.52.1)$$

(3.52) differs significantly from (3.51) in terms of flexibility. With (3.52), the number of product terms increases with increasing iterations. But unlike the previous situation, there is no way to overcome the overflow and underflow behaviour except by lowering the values of λ. This does not serve the purpose because it reduces the computational speed. From the above two cases, it can be inferred that the relaxation operation using β contributes to improved convergence, provided it is controlled suitably.

The use of β, although relevant, along with the iterated form of λ adds computational difficulties to the procedure. This is shown by substituting (3.50.2) and (3.50.3) in (3.50.1) and expressing the result as

$$\hat{f}_j^{z(i)} = \hat{q}^{(i)} \prod_{r=0}^{i} \prod_{s=0}^{i} \hat{f}_{j1}^{z(r,s)} \qquad (3.53)$$

where

$$\hat{f}_{j1}^{z(r,s)} = \exp\left\{ -\lambda^{(r)} \frac{\partial Q^{(s)}}{\partial f_j} \right\} \qquad (3.53.1)$$

The situation is rather complex, since the number of exponential product terms is squared with their magnitude increasing with increasing values of λ. These product estimates at various iterations are scaled and summed up subsequently to form the intermediate estimate given by

$$\tilde{f}_j^{(n)} = \sum_{i=0}^{n} \beta_p^{(i,n)} q^{(i)} \prod_{r=0}^{i} \prod_{s=0}^{i} \tilde{f}_{jl}^{(r,s)} \qquad (3.54)$$

We note that at any given iteration n, for the same number of terms in $\beta_p^{(i,n)}$ as in (3.51.2), by recursive variation of λ, the number of exponential product terms have been squared in (3.54). Therefore, the value of β is expected to be significantly lowered. The complexity of the situation demands a quantitative approach in the determination of λ and β for reliable improvements in the procedure.

3.3. PHASE-III: PARAMETER CONTROL AND CONVERGENCE

The quantitative determination of λ and β, as seen above, and also as reported in the literature, is not straightforward in the procedure. This section starts with the formulation of a set of criteria to be satisfied in the procedure for convergence improvement. Techniques can then be subsequently developed to fulfill these criteria for the procedure. With the parameters satisfying the criteria, the convergence status of the procedure can be analysed.

3.3.1. Criteria for controlling λ and β

The analysis results has shown that β, which appears as a relaxation parameter in the procedure exercises an influence on the exponential argument along with λ, which is an increasing parameter with iterations. Reducing the values of λ is not a solution because it tends to decrease the rate at which the iterations can proceed. Under these conditions the criteria can be stated in their order as follows:

1. The convergence of iterations should be made independent of the variations of λ. By this, the component K_λ of $\overline{\lambda}$ acquires enough freedom to control the rate of convergence.
2. Recalling the discussions of section 3.1.9, $\tilde{\tilde{f}}_j$ in (3.35.2) should be made equivalent to

its true form in (3.7).

3. With the criteria 1 and 2 satisfied, the sufficient condition for iterative convergence given by (3.32) should be satisfied at all iterations and for all pixels.

3.3.2. Controlling techniques

Since the fundamental purpose behind the above criteria is improved convergence, let us start with an evaluation of the first derivative of the iterative function g(x). Rewriting (3.37) gives

$$\underset{f_j}{\vec{f}}^{(i,\,n+1)} = \sum_{i=0}^{n+1} \underset{f_j}{\vec{f}}^{(i,\,n+1)} \tag{3.55}$$

where

$$\underset{f_j}{\vec{f}}^{(i,\,n+1)} = \beta_p^{(i,n+1)} \underset{f_j}{\vec{f}}^{(i)} \tag{3.55.1}$$

The iterative function can be written as

$$g\left(\underset{f_j}{\vec{f}}^{(i,n+1)}\right) = \left(\underset{f_j}{\vec{f}}^{(i,n+1)}\right) = \beta_p^{(i,\,n+1)} \underset{f_j}{\vec{f}}^{(i)} \tag{3.56}$$

Substituting (3.53) and (3.36.1) in (3.56) and expressing the result in terms of the i^{th} and remaining components yields

$$g\left(\underset{f_j}{\vec{f}}^{(i,\,n+1)}\right) = \beta^{(i)} q^{(i)} \underset{f_{j1}}{\vec{f}}^{(i)} \prod_{t=i+1}^{n} \left(1-\beta^{(t)}\right) \prod_{r=0}^{i-1} \prod_{s=0}^{i-1} \underset{f_{j1}}{\vec{f}}^{(r,s)} \tag{3.57}$$

where

$$\underset{f_{j1}}{\vec{f}}^{(i)} = \underset{f_{j1}}{\vec{f}}^{(i,i)} = \exp\left\{ - \lambda^{(i)} \frac{\partial Q^{(i)}}{\partial f_j} \right\} \tag{3.57.1}$$

and

$$\underset{f_{j1}}{\vec{f}}^{(r,s)} = \exp\left\{ - \lambda^{(r)} \frac{\partial Q^{(s)}}{\partial f_j} \right\} \tag{3.57.2}$$

Working out the magnitude of the derivative of (3.57) results in

$$\left| g\left(\tilde{\tilde{f}}_j^{(i,n+1)}\right)\right| = A^t A\, \beta^{(i)} \lambda^{(i)} q^{(i)} \tilde{\tilde{f}}_{j1}^{(i)} \prod_{t=i+1}^{n+1} (1-\beta^{(i)}) \prod_{r=0}^{i-1} \prod_{s=0}^{i-1} \tilde{\tilde{f}}_{j1}^{(r,s)} \qquad (3.58)$$

Expressing the result in terms of $\tilde{\tilde{f}}_j^{(i)}$ using (3.53) yields

$$\left| g\left(\tilde{\tilde{f}}_j^{(i,n+1)}\right)\right| = A^t A\, \beta^{(i)} \lambda^{(i)} q^{(i)} \tilde{\tilde{f}}_j^{(i)} \prod_{t=i+1}^{n+1} (1-\beta^{(i)}) \qquad (3.59)$$

A^t denotes the transpose of the measurement function matrix A such that $|A^tA| \leq 1$. Now, $\beta^{(i)}$ can be chosen to satisfy the various criteria listed above. The first criterion is satisfied by cancelling the variations of λ through β. That is by choosing

$$\beta^{(i)} = \frac{K_\beta}{\lambda^{(i)}} \qquad (3.60)$$

where K_β is a positive constant. The denominator term $\lambda^{(i)}$ is straightforward to calculate from (3.48) as $\lambda^{(i)} = K_\lambda/\bar{Q}^{(i)}$. The numerator term is chosen to be,

$$K_\beta = \frac{1}{\sum_{k=1}^{n} \tilde{\tilde{f}}_k^{(n)}} \qquad (3.61)$$

in order to satisfy the second criterion. Appendix-E shows how the second criterion is satisfied using (3.61). The resulting form of $\beta^{(i)}$ is

$$\beta^{(i)} = \frac{1}{\lambda^{(i)} \sum_{k=1}^{n} \tilde{\tilde{f}}_k^{(i)}} \qquad (3.62)$$

Turning to the third criterion, the sufficient condition to force convergence in the iterative procedure follows directly from the well-known condition (3.32), written as

$$\left| g\left(\tilde{\tilde{f}}_j^{(n)}\right)\right| = \sum_{i=0}^{n} \left| g\left(\tilde{\tilde{f}}_j^{(i,n)}\right)\right| < 1. \qquad (3.63)$$

The right hand side term follows from (3.59), with $\beta^{(i)}$ substituted from (3.62). The result

is:

$$\left| g'\left(\bar{\tilde{f}}_j^{(i,n)}\right) \right| = A^t A \; \frac{\bar{\tilde{f}}_j^{(i)}}{\sum_{k=1}^{n} \bar{\tilde{f}}_k^{(n)}} \prod_{t=i+1}^{n} \left(1-\beta^{(t)}\right) \tag{3.64}$$

Substituting (3.64) in (3.63) and using the relation (3.55), we get

$$\left| g'\left(\bar{\tilde{f}}_j\right)^{(n)} \right| = A^t A \frac{\bar{\tilde{f}}_{jd}^{(n)}}{\sum_{k=1}^{n} \bar{\tilde{f}}_k^{(n)}} \tag{3.65}$$

where -

$$\bar{\tilde{f}}_{jd}^{(n)} = \sum_{i=0}^{n} \beta_{p1}^{(i,n)} \bar{\tilde{f}}_j^{(i)} \tag{3.65.1}$$

In (3.65.1), $\beta_{p1}^{(i)}$ is the same as (3.36.1) but with $\beta^{(i)} = 1$. Comparing (3.65) and (3.63), we can see that

$$\left| g'\left(\bar{\tilde{f}}_j^{(n)}\right) \right| < 1 \text{ if } \bar{\tilde{f}}_{jd}^{(n)} < \sum_{k=1}^{n} \bar{\tilde{f}}_k^{(n)}. \tag{3.66}$$

In writing (3.66), the condition $|A^tA| \leq 1$ has been used. Also for a given $\beta^{(i)}$, $\tilde{\bar{\tilde{f}}}_j^{(n)} < \bar{\tilde{f}}_{jd}^{(n)} \leq \bar{\tilde{f}}_j^{(n)}$, since $\beta^{(i)} \leq 1$ in obtaining $\bar{\tilde{f}}_j^{(n)}$ whereas $\beta^{(i)} = 1$ for $\bar{\tilde{f}}_{jd}^{(n)}$ and there is no scaling at all in the case of $\tilde{\bar{\tilde{f}}}_j^{(n)}$. The condition (3.66) can then be restated as

$$\left| g'\left(\bar{\tilde{f}}_j^{(n)}\right) \right| < 1 \text{ if } \bar{\tilde{f}}_j^{(n)} < \sum_{k=1}^{n} \bar{\tilde{f}}_k^{(n)}. \tag{3.67}$$

The existence and finiteness of $\tilde{\bar{\tilde{f}}}_j^{(i)}$, however is based on the numerical stability considerations which is the subject matter of the following section.

3.3.3. Convergence conditions in the procedure

The condition (3.67) is indeed consistent with the behaviour of the unaveraged estimate, which is already known to be a factor causing numerical instabilities. Rewriting

the unaveraged estimate from (3.50.1)

$$\tilde{f}_j^{(i)} = \tilde{q}^{(i)} \exp\left\{ -\bar{\lambda}^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \right\} \tag{3.68}$$

The value of $\tilde{f}_j^{(i)}$ at an iteration i depends upon the value of the exponential argument, which in turn is determined by the values of $\bar{\lambda}$ and $\partial\bar{Q}/\partial f_j$. Dividing (3.68) by $\sum_k \tilde{f}_k^{(i)}$, at any iteration i, the condition in (3.67) can also be expressed as

$$\frac{\tilde{q}^{(i)}}{\sum_{k=1}^{n} \tilde{f}_k} \exp\left\{ -\bar{\lambda}^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \right\} < 1. \tag{3.69}$$

Expressing the condition in terms of the exponential argument we obtain

$$x_j^{(i)} < \log \left(\frac{\sum_{k=1}^{n} \tilde{f}_k}{\tilde{q}^{(i)}} \right) \tag{3.70}$$

or simply using (3.2.2):

$$x_j^{(i)} < S(\tilde{\mathbf{f}}) \text{ for } j = 1, \ldots\ldots, n \tag{3.70.1}$$

where

$$x_j^{(i)} = -\bar{\lambda}^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \tag{3.70.2}$$

and

$$S(\tilde{\mathbf{f}}) = -\sum_{i=1}^{n} \frac{\tilde{f}_i}{\sum_{k=1}^{n} \tilde{f}_k} \log \frac{\tilde{f}_i}{\sum_{k=1}^{n} \tilde{f}_k} \tag{3.70.3}$$

(3.70), (3.70.1) are the restatements of the condition (3.67) in different forms. (3.70.1) shows that for convergence to take place, the exponential argument should be less than the entropy of the estimated probability distribution.

Let us consider the actual situation encountered by the procedure, that is $\bar{\tilde{f}}_j^{(i)}$

resulting from the use of $\beta^{(i)}$. From equation (3.35), the exponential argument can be written as

$$x_j^{(i)} = \log\left(\frac{\tilde{f}_j^{(i)} - (1-\beta^{(i)})\tilde{f}_j^{(i-1)}}{\beta^{(i)}\tilde{q}^{(i)}}\right). \tag{3.71}$$

Denoting the numerator term by $\tilde{f}_{js}^{(i)}$ and substituting for $\beta^{(i)}$ in the denominator, (3.71) becomes

$$x_j^{(i)} = \log\left(\frac{K\lambda \tilde{f}_{js}^{(i)} \sum_{k=1}^{n} \tilde{f}_k^{(i)}}{Q^{(i)}\tilde{q}^{(i)}}\right). \tag{3.72}$$

The term $\tilde{f}_{js}^{(i)}$ is given by

$$\tilde{f}_{js}^{(i)} = \tilde{f}_j^{(i)} - (1-\beta^{(i)})\tilde{f}_j^{(i-1)}. \tag{3.72.1}$$

Expanding the logarithmic term in (3.72) and using (3.48) we get

$$x_j^{(i)} = \log \tilde{f}_{js}^{(i)} + \log \lambda^{(i)} + S(\tilde{f}^{(i)}). \tag{3.73}$$

Equivalently we obtain

$$x_j^{(i)} = \log \tilde{f}_j^{(i)} + \log \beta^{(i)} + \log \lambda^{(i)} + S(\tilde{f}^{(i)}). \tag{3.73.1}$$

The variations of the term $\tilde{f}_{js}^{(i)}$ depend on how the estimates $\tilde{f}_j^{(i)}$ vary with the iterations. In general, we can write

$$\tilde{f}_{js}^{(i)} > 0 \text{ if } \tilde{f}_j^{i} > \tilde{f}_j^{i-1}. \tag{3.74.1}$$

$$\tilde{f}_{js}^{(i)} \gtrless 0 \text{ if } \tilde{f}_j^{i} < \tilde{f}_j^{i-1}. \tag{3.74.2}$$

Combining the conditions in (3.74.1) and (3.74.2) with (3.73.1), we see that practically, the exponential argument exceeds the entropy of the estimated probability distribution with increasing values of λ. The condition (3.70.1) is satisfied only when

$$\lambda^{(i)} \bar{f}_{js}^{(i)} \leq 1 \tag{3.75}$$

From (3.72.1), we can see that (3.75) is satisfied when

$$\beta^{(i)} = 1 \tag{3.76}$$

and when $\beta^{(i)} \rightarrow 0$ it is true that

$$\bar{f}_{js}^{(i)} \rightarrow 0. \tag{3.77}$$

The first condition is satisfied when i = 0. This corresponds to the unconstrained maximization point. Using equation (3.35), the equivalent condition of (3.77) can be written as

$$\bar{f}_j^{(i)} \rightarrow \bar{f}_j^{(i-1)} \tag{3.78}$$

When the entropy of the estimate becomes maximum, the condition (3.78) becomes active in the procedure, which is the convergence condition. Depending on the parameter λ, β regulates the process for convergence. The convergence processes based on (3.77) in general, have a hidden implication. (3.77) or (3.78) do not imply anything at all about the absolute magnitude of the resulting estimates. This means that the indirect scaling information in the data constraint is lost in the convergence process. One of the test studies in the following chapter demonstrates this situation. The scaling information can be provided using an external constraint compatible with (3.77). The information on the maximum possible limit for the estimate values fulfills this condition. For images, the maximum possible value is the maximum level on the scale used for the input image scaling. This constraint is referred to as the **upper-bound constraint** in the discussions that follow. By placing this constraint in the procedure, the convergence process is constrained by the scaling information. A measure based on (3.78) for all pixels can sense the convergence condition of the procedure externally. Such a measure is indeed the standard convergence measure, written as

$$\gamma^{(n)} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{f}_j^{(n)} - \hat{f}_j^{(n-1)}\right)^2} \qquad (3.79)$$

$\gamma^{(n)}$ is expected to be close to 0 for the result to be a true solution in an iterative equation.

3. 4. CONVERGENCE MECHANISM OF THE PARAMETER RELAXATION PROCEDURE

The direction of analysis followed in this chapter has led to the development of a simple and computationally feasible (stable and convergent) procedure to obtain a *ME* solution, which is the primary aim of this chapter. The analysis results of the various intermediate stages, when viewed in a global perspective give rise to a more meaningful interpretation of the procedure as discussed below.

The information processing activities underlying the computational scheme are best explained as interactions between a system $\mathbf{S} = \{\log \hat{f}_j\} ; j = 1,.....,n\}$ and its environment, which in the present case is an algorithm based on the relaxation scheme. The algorithm influences the system by controlling its logarithmic states $\mathbf{s}_j$ thereby changing the 'behaviour' of the system. In the discussions below, it implies that $\mathbf{s}_i \in \mathbf{S}$ for all i's. As the equation (3.73) shows, in the present study, the behaviour of each state can be conveniently described by the exponential argument $x_j^{(i)}$ for $j = 1, ..., n$.

Initially, the state of the system is stable. This system is characterized by no external constraints and has an entropy of $\log n$. The initial logarithmic state of the system (equation 3.11) is given by

$$\mathbf{s}_0 \leftarrow \hat{\mathbf{f}}^{(0)} = \left\{\hat{f}_j^{(0)}\right\} \quad j = 1, \cdots\cdots n. \qquad (3.80.1)$$

Combining the inequality (3.70.1) with (3.68), the behaviour of the system can be

described as

$$\mathbf{B}(\mathbf{s}_0) < S(\hat{\mathbf{f}})^{(0)}_{\hat{q}_a^{(0)}} \qquad (3.80.2)$$

$S(\hat{\mathbf{f}})^{(0)}$ is the entropy of the system calculated using (3.70.3) for the initial state given by (3.11). $\hat{q}_a^{(0)}$ is the logaritmic average obtained by taking the logarithm of $q(\hat{\mathbf{f}})$. $\mathbf{B}$ simply denotes a numerical scale for the behaviour measure. When an external constraint (data) acts upon the system, its behaviour changes. Depending on the strength of the external constraint parameters λ and σ^2, the states get changed to

$$\tilde{\mathbf{s}}_i \leftarrow \tilde{\mathbf{f}}^{z(i)} \qquad (3.81.1)$$

with $\tilde{\mathbf{f}}^{(i)}$ given by (3.35.2). The behaviour of these states is based on (3.73) written as

$$\tilde{\mathbf{s}}_i \leftarrow \log f_j^{z(i)} = \left(\log \tilde{f}_{js}^{\,z(i)} + \log \lambda^{(i)} + S(\tilde{\mathbf{f}})^{(i)} \right)_{\tilde{q}_a^{z(i)}} \qquad (3.81.2)$$

which implies that

$$\mathbf{B}(\tilde{\mathbf{s}}_i) > S(\tilde{\mathbf{f}})^{(i)}_{\tilde{q}_a^{z(i)}} \qquad (3.81.3)$$

$S(\tilde{\mathbf{f}})^{(i)}$ is calculated using (3.70.3) using (3.35.2). As the strength of the constraint λ increases, the beaviour of the system becomes unbounded. The system is controlled by the algorithm using the relaxation parameter $\beta^{(i)}$. This relaxation process modifies the states (equation 3.35) to

$$\bar{\mathbf{s}}_i \leftarrow \bar{\mathbf{f}}^{z(i)} \qquad (3.82.1)$$

The behaviour (equation 3.73.1) changes to

$$\bar{\mathbf{s}}_i \leftarrow \left\{ \log \bar{f}_j^{\,z(i)} \right\} = \left\{ \left(\bar{f}_j^{\,z(i)} + \log \beta^{(i)} + \log \lambda^{(i)} + S(\bar{\mathbf{f}})^{(i)} \right)_{\bar{q}_a^{z(i)}} \right\} \qquad (3.82.2)$$

From the behaviour of the parameters discussed in connection with equation (3.73), (3.82.2) implies that

$$\mathbf{B}(\bar{\mathbf{s}}^{(i)}) \gtrless S(\bar{\mathbf{f}})^{(i)}_{\bar{q}_a^{(i)}} \qquad (3.82.3)$$

$S(\bar{\mathbf{f}})^{(i)}$ is given by (3.70.3). The states can be effectively relieved from the influence of λ

by controlling β suitably as in (3.62). This relaxation process, when repeated, refines the behaviour of the system successively. When entropy becomes maximum, say $S_m(\bar{\hat{f}})$, which happens when the solution equation (3.8) is satisfied, the condition (3.70.1) becomes

$$x_j^{(i)} \leq S_m(\bar{\hat{f}}) \text{ for } i > m \text{ and } j = 1,\ldots,n. \qquad (3.83)$$

At this stage m, the system begins to obey the condition

$$B(\bar{s}^{(m)}) \leq S(\bar{\hat{f}})^{(m)} q_a^{(m)}. \qquad (3.83.1)$$

(3.83.1) happens when the entropy of the system is close to its true maximum (equation 3.70.1), which is the convergence condition. The corresponding state of the system is the required solution, denoted as

$$\bar{s} \leftarrow \bar{\hat{f}}^{(m)}. \qquad (3.83.2)$$

The scaling of (3.83.2) will be arbitrary, since the relaxation process has taken place without the scaling information. However, with an external upper bound constraint, the process can be controlled depending upon the scaling requirements, so that

$$\bar{s}_{sm} \leftarrow \bar{\hat{f}}^{(sm)} \qquad (3.84)$$

The key activity behind convergence is the relaxation process with associated control mechanisms. Simulated annealing techniques, the currently popular techniques in solving complex constrained optimization problems [27], [72], [73] are reported to have a similar mechanism for successful convergence. The technique is discussed in chapter 7. Reports also indicate that the current technological developments of VLSI array processors are well suited to simulate the algorithmic activities of a relaxation process [76]. The discussions of these issues are also covered in chapter 7. In the next chapter, the behaviour characteristics of the relaxation process and its influence on the states of the system are tested and studied using practical examples. The subject matter is approached from a functional point of view to avoid the abstractions associated with the systems perspective.

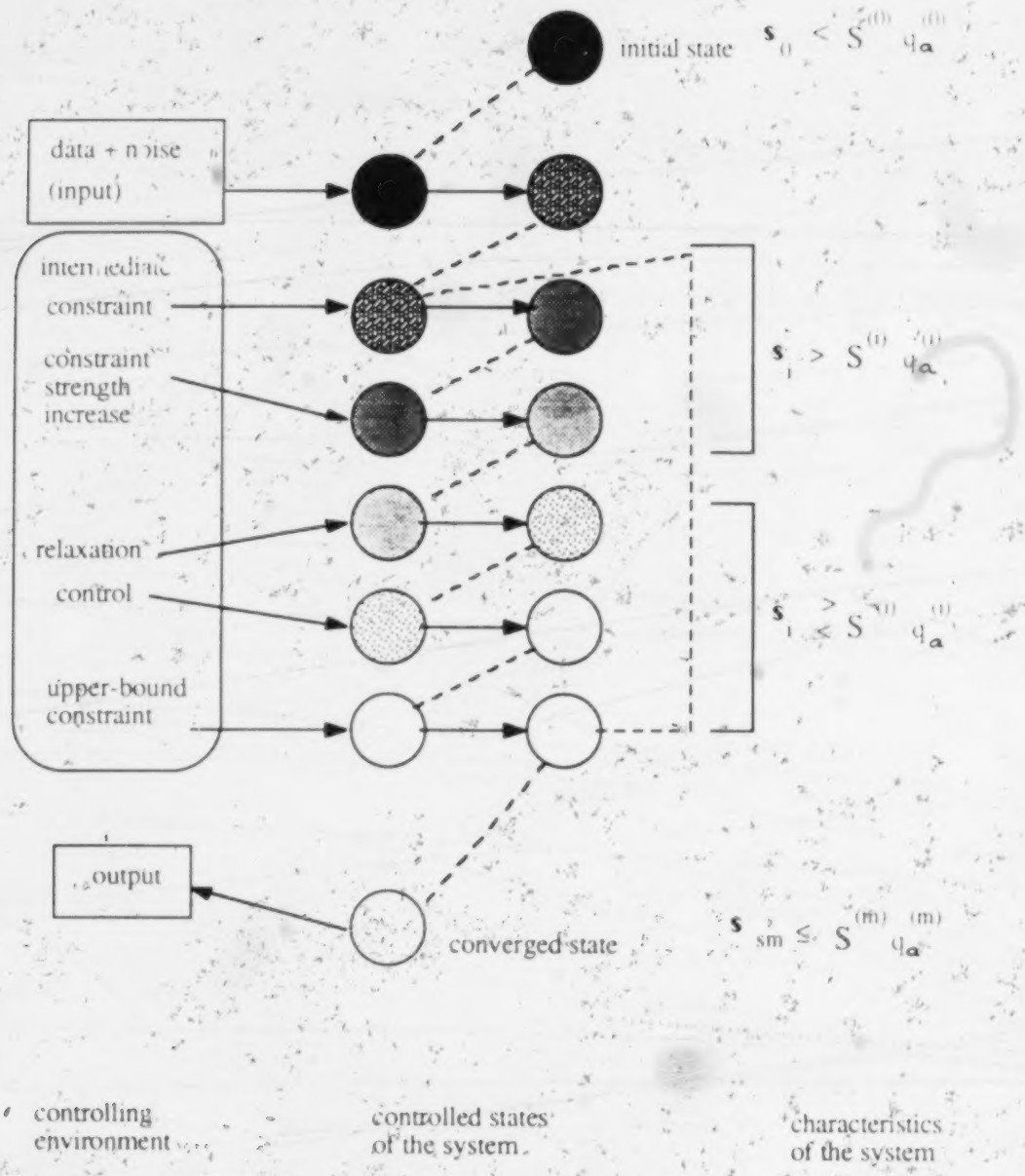


Fig.3 Mechanism of the parameter relaxation procedure.

CHAPTER 4

NUMERICAL AND FUNCTIONAL BEHAVIOUR OF THE RELAXATION SCHEME -TEST STUDIES

The solution scheme developed for the *ME* method in chapter 3 shows two different directions for test studies: the study of its dynamic behaviour characteristics and functional merits. The two aspects are explored in this chapter by conducting test studies on practical examples. The characteristics of the scheme are evaluated analytically and compared with the test results. The functional merits are evaluated in inverse problems. Natural and computer generated images serve as practical examples for all the test studies. Results are discussed in two phases: Phase-I deals with studies on the behaviour characteristics and phase-II with applications.

4.1. TEST OBJECTIVES AND SET-UP

The tests all involve applying the parameter relaxation scheme on practical images using an algorithm. This section begins with an account on the choice of images and the algorithm followed for the test studies.

4.1.1. Choice of images and algorithms

For the study under discussion, the type of images are from the class (3.1) that need inversion of some kind. For test studies, these images are obtained in two stages: In the first stage, 'source images' are acquired from external sources and stored in digital form using an image acquisition or special camera equipment systems. Exceptions to the camera acquistion step are computer generated source images. In the second stage, the 'input images' are generated from the source images using image processing techniques. The

discussions on the generation of the input images are postponed to the phase-II section. For phase-I, images are ensured to be in the class (3.1) and their exact type is not a matter of concern.

When selecting algorithms, the ease of programming and tractability are two major factors to be considered since extensive trial runs are needed for test studies. The time complexity of the sequential algorithms in the application class is as high as $O(n^2)$, where n is the total number of pixels in an image, or the image size. To reduce the computational time demands, images of dimensions 64 x 64 have been considered. Also, the results and conclusions obtained from the test studies using the sequential algorithm are expected to serve as reference for further developments in chapter 7. The sequential algorithm followed for the test studies is outlined below.

4.1.2. The test algorithm

An algorithm by definition is characterized by its input, output, finiteness definiteness and efficiency [40]. The basic specifications of the algorithm used in this study are given below:

Input: The measurement function A and the 2-D set of numbers d_j, representing the measured data or image (finite and bounded) are the inputs to the algorithm. The numbers d_j could be positive or negative and their type could be integer, real or complex depending on the type of application. For example, in the reconstruction of Fourier synthesis data in radio astronomy, the input data are mostly complex [12], [13]. If the input is noisy, then d_j will be positive or negative.

Output : The *ME* estimate of the source image distribution (2-D set of numbers) $\hat{f}_j$, $j = 1, \ldots, n$.

Algorithm: All the test studies make use of an iterative sequential algorithm shown in Fig. 4.1. It is represented as a set of computational steps, each step charaterized by its own input(s) and output(s). The input(s) and output(s) are either a single parameter or 2-D arrays(1-D if the image is stored in a vector format). 2-D arrays are denoted by outlined characters. The algorithm, for its implementation requires three 2-D image data arrays for the output, input and an additional one to hold the previous output. There are four phases in the algorithm: reading of the inputs and initialization constitute the first phase. The data processing in the second phase consists of calculating $\bar{Q}$ and $\partial\bar{Q}/\partial f$, (data space calculations) followed by the estimation of parameters (parameter space calculations). The results of these steps are combined to determine the current estimate using (3.35). The convergence condition is checked in the third phase. The results (estimate images and parameters) of different stages are output in the fourth phase.

Finiteness: Finiteness of an algorithm is assessed by the feasibility of its termination conditions. The termination conditions are determined from the iterative convergence of the procedure measured using

$$\gamma^{(n)} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\bar{f}}_j^{(n)} - \hat{\bar{f}}_j^{(n-1)}\right)^2} \qquad (4.1)$$

$\gamma^{(n)}$ becomes close to 0, upon convergence. The feasibility of the convergence condition in the algorithm follows directly from the derivation results of section 3.3.2.

4.2. PHASE - I: NUMERICAL BEHAVIOUR CHARACTERISTICS

The algorithm explained above was run on six different images shown in Fig 4.2 (a)-(f) all from the class (3.1). The inputs are identified as input-1, input-2, etc, as shown in the captions. The inputs are of different types of data (positive, negative, real, complex) and also correspond to different forms of $\partial\bar{Q}/\partial f$ in the algorithm. Given the inputs, the

control parameters K_λ, $\hat{q}$, $\bar{\lambda}$ and β are determined by the algorithm iteratively. Since the algorithm follows a set of criteria to determine the parameters, the behaviour of the algorithm is predictable mathematically. Experimentally, the characteristics can be determined by recording the variations of all the relevant parameters $\bar{Q}$, $\bar{\lambda}$, β, $\hat{q}$, S and γ with iterations under different input and operating conditions. These characteristics can subsequently be compared against the predicted ones. The derivation of the scheme in chapter 3 shows a definite purpose for each of the control parameters in the algorithm. Based on this understanding, the parameter $\bar{\lambda}$ is qualified as the 'constraint strength' parameter and β as the 'relaxation parameter' in the following discussions. These two parameters are the design parameters, whereas $\hat{q}$ is the 'scale' parameter is part of the scheme itself. In the discussions that follow, $\bar{Q}$ and $\bar{\lambda}$ will be referred to as Q and λ for convenience.

4.2.1. Predicted Vs recorded behaviour of the algorithm

As indicated by the equation (3.41), Q is an error measure for the intermediate estimates. This means that the algorithm evaluates the functional distance between the intermediate and the true (solution) estimates from the value of Q in each iteration. Based on this measure, the algorithm estimates the control parameters λ and β to determine the next estimate. In addition, the algorithm also makes use of the information derived from the previous intermediate estimates using the scale paramter q in the estimation process. The control mechanism of the algorithm is illustrated in Fig. 4.3.

4.2.1.1. Constraint strength parameter λ

In the closed loop of control in Fig. 4.3, the variations λ can be determined relative to variations of Q. It is obtained from (3.48) as

$$\frac{d\,\bar{\lambda}^{(n,i)}}{d\bar{Q}^{(i)}} = -\frac{K_\lambda}{\bar{Q}^{2(i)}} \qquad (4.2)$$

The notation $\lambda^{(n,i)}$ conveys the meaning that the value of λ at iteration n depends upon its value at n, as well as the other iterations. In the discussions that follow, the term $\lambda^{(n,i)}$ will be simply referred to as λ. The right hand side term of (4.2) is strictly a negative quantity, meaning that the variations of λ and Q are in opposite directions. Since the variation of λ is known to be strictly positive (increasing in values with increasing iterations), its influence on the intermediate estimate takes place in the direction of decreasing Q. The decrease in the values of Q as noted in chapter 3, is an indication of progress towards convergence. Any external constraint imposed by the algorithm on the intermediate estimates, such as the one obtained by imposing the upper bound constraint may change the slope of Q. This in turn changes the slope of λ but not its direction of variations. Therefore, the algorithm controls λ such that it is strictly increasing with iterations but its slope variations are controlled depending on the variations of Q. The results of running the algorithm on the input images illustrate this behaviour of λ over 20 iterations in Fig 4.4 (c) and (d). The variations of Q are shown in Fig 4. 4 (a) and (b). The values of Q varied widely between 6000 and 15000 for the different inputs. These values have been converted to a common scale in plotting the graph. For all the input cases, the slopes of λ change at a particular iteration and as will be seen shortly, are due to the influence of the upper bound constraint on the intermediate estimates.

4.2.1.2. Relaxation parameter β

The purpose of using β in the algorithm is to reduce the strength of the influence of the data constraint (increasing λ) on the intermediate estimates. As discussed in chapter 3, it is done by using β in a relaxation scheme (equation 3.35) with its variations in a compensating direction to that of λ. The expected relative variations of β can be found from

(3.62) as

$$\frac{d\,\beta^{(n)}}{d\,\bar{Q}^{(i)}} = \frac{1}{K\lambda \sum_{k=1}^{n} \bar{f}_k^{(i)}} \qquad (4.3)$$

Since the intermediate estimates are all strictly positive, the denominator of (4.3) is a positive quantity, which means that the algorithm controls β by following closely the error variations. Fig 4.4 (e) and (f) shows the experimental results for illustration. Here also, the values are calculated on a common scale of 0-1 to accommodate all the six cases in a single plot. As can be seen in the figure, when Q tends to change its directions, the algorithm changes the course of β accordingly.

4.2.1.3. Scale parameter q

The variation of q, as the equation (3.35.3) indicates, is a function of the intermediate estimates. Its variation is governed by the relation

$$\frac{\partial\, q^{(n)}}{\partial \bar{f}_j^{(n)}} = \frac{q^{(n)}}{\sum_{k=1}^{n} \bar{f}_k^{(n)}} \left(1 + \left(\bar{\hat{p}}_{jn}^{(n)} \log \bar{f}_j^{(n)}\right)\right) \qquad (4.4)$$

where

$$\bar{\hat{p}}_{jn}^{(n)} = 1 - \bar{\hat{p}}_j^{(n)} \qquad (4.4.1)$$

Unlike the parameters λ and β, the parameter q is an 'implicit' control parameter in the algorithm. Equation (4.4) shows that the direction of variation of q for each pixel is based on the sign of the term $(1+\bar{\hat{p}}_{jn}^{(n)} \log \bar{\hat{f}}_j^{(n)})$. It implies that the variations are strongly influenced by the initial estimate of the algorithm. Hence, unlike λ and β, the relative behaviour of q does not exhibit a fixed pattern in the algorithm. However, upon convergence, irrespective of the other parameter conditions, q becomes constant in the algorithm. This expectation follows directly from the convergence condition in (3.78). For the algorithm under study,

the initial estimates are all ≥ exp(-1), which means that the parenthesis terms are ≥0 initially. It sets up a uniform initial base for a monotonic increase of the intermediate estimates. Hence $q^{(n)}$ is expected to increase steadily up to the convergence point. The observed behaviour for the inputs shown in Fig. 4.4 (g) and (h) supports this view point.

4.2.1.4. Entropy S

Along the same lines, the entropy variations can be determined from (3.70.3) and written as

$$\frac{\partial S(\bar{\bar{f}}^{(n)})}{\partial \bar{\bar{f}}_j^{(n)}} = \frac{1-\bar{\bar{p}}_j^{(n)}}{\sum_{k=1}^{n} \bar{\bar{f}}_k^{(n)}} \left(1+\log \bar{\bar{p}}_j^{(n)}\right) \qquad (4.5)$$

The variation of entropy is determined by the sign of $(1+\log \bar{\bar{p}}_j^{(n)})$, which in turn is a function of the initial and the intermediate estimates. Its relative behaviour along with q, is difficult to predict. But once again, with uniform initial estimate and monotonic variations of the intermediate estimates, its behaviour is predictable in an intuitive sense. $S(\bar{\bar{f}})$ being a measure of uncertainty, is expected to decrease during the 'active phase' of the iterations. Active phase is the phase between the initial stage and the beginning of the convergence. During this phase, because of the steady influence of the external constraints, the degree of uncertainty decreases in the intermediate estimates and hence the entropy decreases. Once convergence point is reached, the entropy variations become consistently positive because of its maximum conditions in the algorithm. The observed variations in Fig.4.4 (i) and (j) indeed show this behaviour for all the inputs. However, these arguments cannot be generalized for all types of initial estimates.

4.2.1.5. Convergence

In all the six test cases, there are no signs of overflow or underflow in the behaviour of the algorithm, which is an indication of numerical stability. The variations of

the convergence measure γ shown in Fig 4.4(k) and (l) are indicative of this improvement. In the termination phase, the constraint factors have no effect on the intermediate estimates, which is the characteristics of a convergent state. In particular, it may be noted that a linear increase in λ during termination phase has not influenced the state of convergence. To study the influence of the upper bound constraint on convergence, the following test was performed.

4.2.2. Removal of upper-bound constraint: a test for convergence

From the above studies it was understood that the upper-bound constraint imposed by the algorithm on the intermediate estimates has an influence on the variations of Q, λ and β. This is because, the upper-bound constraint, as discussed in chapter 3 incorporates the scaling information of the input data in the estimation process. The constraint however does not influence the convergence of iterations. To test this, the algorithm is studied for its numerical behaviour in the absence of the upper bound constraint. According to the analytical results in chapter 3, if the upper bound constraint is removed, the magnitude of the intermediate estimates will begin to increase at a point when the intermediate estimates exceed the range of input scaling. As a result, Q will increase, the rate of increase being dependent upon the control parameters. This scale change in the intermediate estimates will obviously result in the sudden changes of entropy. Despite these discrepancies, the algorithm is expected to control the iterations for convergence because, as understood earlier, the convergence process does not depend upon the scaling and follows naturally when the entropy reaches its maximum limits, whatever it is. Specifically, the algorithm is expected to respond with λ still increasing but with a much smaller slope. β is expected to be modified for its value to the new variations of Q in order to maintain numerical stability. With this understanding, the procedure was put to the following test: The algorithm was run on an arbitrarily chosen example, in this case, input-3 without imposing the

upper-bound constraint. The resulting characteristics are shown in Fig 4.5 (a)-4.5(l). For a clear representation, the plots are shown separately for the active and the termination phases with iteration 10 separating the two phases. In the active-phase Q begins to decrease in a normal manner (as in 4.4 (c)-4.4 (d)). At iteration 11, when the parameter settings of the procedure are ready for scaling (the input image scaling is 0-255), there is no active mechanism to define the upper limit. The iterations therefore continue to respond to increasing λ without an upper-limit, which causes an apparent outburst of numbers due to exponential operation. Q as expected, has a wide swing in its variations (1300 times higher than the normal) and λ changes its slope. The entropy shows these changes by a sudden decrease in its course. The interesting factor here is that β adjusts to this new situation by considerably raising its value (880 times larger than the normal case). The fact that there are no integer overflows despite the exponential growth of the estimates is an indication of the numerical stability of the procedure. Further evidence for stability comes from the convergence variations shown in Fig4.5 (l). The characteristics of the convergence are seen in the termination phase, but the value of the convergence measure in the termination phase was found higher, close to 10 (is not clear in the plot due to the wide range of scaling).

4.2.3. Influence of K_λ on convergence speed

The derivation results of chapter 3 show that the choice of K_λ in λ is not very critical to the functioning of the procedure and can be adjusted for faster convergence. To test this, the algorithm was run on an arbtrarily chosen image, once again input-3 with three different values of K_λ (30, 70 ,100). The resulting Q and the convergence variations are shown in Fig 4.6. The lowest K_λ (30) has significantly lowered the termination value of Q but has not converged within the observed window of 20 iterations. As expected, with

increasing values of K_λ, the number of iterations needed for convergence decreases. However, this has the effect of increasing the limiting values of Q slightly above its settled value in the active phase. In practice, these variations do not seem to produce a noticeable effect in the final results. With very high value of K_λ (a quantitative limit is not available), the behaviour of the procedure was found to be normal but its application performance was not satisfactory. The convergence value of Q was found to be much higher than its settled value in the active-phase and the resulting output image appeared distorted. This situation can be overcome by working out a reasonable upper limit for K_λ. The upper-limit is expected to be a function of the image size and the total intensity of the input image.

4.2.4. Error performance

Error analysis results in chapter 3 have shown that the intermediate error terms follow a Gaussian distribution regardless of the presence or the absence of noise error terms in the data (sections 3.1.6 and 3.2.2). This feature has also been noted and reported as an observation in the studies related to simulated annealing techniques [27, page 269]. To substantiate this result, the noiseless but convolved data (input-1) was considered as input to the algorithm. The intermediate error image formed by the quantities

$$\hat{\zeta}_j^{(n)} = d_j - \sum_{k=1}^{n} f_k A_{jk};\ j = 1, \cdots\cdots, m \qquad (4.6)$$

was recorded for different iterations. The histograms of this image at different iterations were estimated (after removing the mean) and shown as a plot in Fig 4.7(a). The results indeed show that the intermediate error terms have formed a Gaussian distribution in the procedure. Also, the decreasing widths of the histogram shows that the intermediate error variance decreases with increasing iterations.

In presence of noise, the derivation results showed that the noise variance

automatically gets reduced in the procedure through the intermediate variance. To test this, the algorithm was run on the noisy image example in Fig 4.2 (input-4). The variance of the noise terms in this example is 368, with the signal to noise ratio of 9.17dB. The variance parameter was not used by the algorithm but used for testing. To show that the procedure indeed reduces noise, a histogram was calculated from the residual quantities given by

$$n_j = \frac{1}{\sigma}\left(d_j - \sum_{k=1}^{n} A_{jk}\hat{f}_k\right) \quad \text{for } j = 1,..n \qquad (4.7)$$

where σ is the standard deviation of the Gaussian noise terms. If the procedure has fitted the noise well, the histogram formed out of the residual quantities should follow a unit Gaussian histogram. These plots for the example under consideration are shown in Fig. 4.7(b) for different iterations. As expected, the width approaches unity with increasing iterations. From the calculations of the area of the histogram, it was also observed that all the residual terms lie within the unit width. The noise performance of the procedure will be further discussed under phase-II section.

4.3. PHASE - II: FUNCTIONAL PERFORMANCE

The objective of this section is to explore the usefulness of the procedure and the *ME* method in practical image processing applications. The lack of an objective measure for images leaves an element of ambiguity in the usage of the terms 'application' and 'usefulness'. This section begins with an understanding of these terms as applied to the work reported in this thesis.

4.3.1. Application perspective

Image processing and machine vision studies are branches of visual science that deal with variety of images encountered in different application fields of science and

engineering. For most of these applications, human visual responses are the only known reliable sources [28], [77], [78]. Indeed the development of an objective measure for image quality itself depends on the psycho-physical understanding of human visual responses. This is because the 'objectiveness' of the measure is expected to match with the judgments of human eyes. Because of this fundamental difficulty, true objective evaluation schemes have not yet evolved for image processing applications. Objectivity at other levels, for example, formalizing a set of test patterns for a given application purpose, contributes to a stable and a systematic evaluation of processing methods and techniques. But the progress in this direction is hampered by the interpretation difficulties of the evaluation process. However, the issue is not considered an obstacle because subjective improvements are one of the application requirements for many image processing fields. Subjectively, images are often assessed for their quality based on the degree of sharpness and details [79], [77], [28]. Reports also indicate objective evaluation measures such as least mean square error but their usage is limited to selected cases. In this study, the application results are evaluated on the subjective as well as objective grounds. Subjective evaluations are however approached with the viewpoint of further developments.

4.3.2. Advantages of the *ME* method

The non-negative nature of the resulting output is one of the well-recognized features of the *ME* method in image processing applications. Further, on the grounds of objectivity, in a similar sense to formalization, the *ME* method has a unique advantage. The objectivity is ensured in the *ME* solutions since the method incorporates all the available information with maximum uncertainty [46]. Extra details cannot appear in a *ME* solution unless there is evidence for it in the input data [18]. This increases the user's degree of confidence on the extra details (if any) revealed by the method. This practical implication of the *ME* method stems directly from its foundations. Further, in the study of inverse problems as a class, the general basis of the method is of significant advantage for

algorithm implementation.

4.3.3. Choice of source images

As mentioned in section 4.1.1. of this chapter, images used for test studies are two-stage products. Source images acquired in the first stage are processed in the second stage to generate input images suitable for the application under study. The source images chosen for this study are shown in Fig. 4.8(a), (b) and (c). The eye image is from a natural object and the text image is from a man-made object. The third image is a simple two-level computer generated pattern. All images are monochrome images with 8-bit resolution. A close examination of the checker board in Fig. 4.8(c) will show discontinuities along the pattern. These 'testgaps' have been introduced while generating the pattern itself for the purpose of testing. The varying types of these source images can be seen from their frequency domain representations in Fig. 4.8(d), (e) and (f). The reasons for choosing an eye image for this study are two fold: First, the general features of an eye are known a prioi and hence provide a certain degree of objectivity basis for interpretation. The second reason is that some of its features, pupil and its reflection for example, aid the visual evaluation of deconvolution performance of the algorithm.

4.3.4. Generation of input images

The input images to the test algorithm are the degraded or the transformed images, obtained from the source images. One of the commonly encountered degradation is blurring and is obtained by convolving the true image with filter or window functions. The amount of degradation in the input image depends upon the type of filter chosen. In this study, a 3 x 3 'normal' and a 7 x 7 'worst case' filters are chosen. The normal filter is a Gaussian window. The other filter is called worst case filter because, spatially, it has its 24 border elements unity, the rest all being zeroes. The transfer function of this filter has many 0's in

its pass band, which is an ideal test setting for inversion, specifically in restoration. The transfer function of these filters are shown in Figs 4.9(a) and (b).

The input images generated using the two filters are shown in Fig.4.10(a), (b) and (c) along with their respective Fourier spectra in (d), (e) and (f). The eye image is the result of convolving the source image in 4.8(a) with the 3 x 3 filter. The other two images have been generated by convolving the high frequency source images in Fig 4.8(b) and (c) with the worst case window filter. Comparison of the Fourier transforms in Fig. 4.10 with Fig.4.8 indicates a high degree of loss of information for the images (b) and (c). The result of convolving the eye image with a 3 x 3 gaussian window, on the other hand has a minimal blurring. It is in fact hard to consider it as a degraded case, because, visually, it appears smoother compared to the source image. The image still qualifies as a test case for the studies, because, whatever may be the input image, it is the *ME* estimate that is under investigation in the present context. With respect to the application class, these are all examples of the case

$$d_j = \sum_{k=1}^{n} f_k A_{jk};\ j = 1, \ldots\ldots, n. \qquad (4.8)$$

The ill-posedness is mainly due to the the singular nature of A. Let us understand these examples as category-I.

The second category of problems include the inevitable source of degradation, the noise terms. In generating noisy input images, the noise terms (zero mean Gaussian noise) are first generated and added (signal independent terms) to the blurred image. Fig. 4.10(g) is one such image with the noise terms of standard deviation of 19.13 added to the blurred eye image in 4.10(a). Its frequency spectrum is shown in Fig. 4.10(h). The signal to noise ratio of the input image is 9.17dB. The image falls under the class

$$d_j = \sum_{k=1}^{n} f_k A_{jk} + e_j;\ j = 1, \ldots\ldots, n. \qquad (4.9)$$

where the e_j's reprent the Gaussian noise error terms. Here, the ill-conditioned nature of the problem is due to the singularity as well as the randomness in the input data.

The third category is a simple example of reconstruction obtained by uniformly skipping the image samples. The input image in Fig. 4.10 (i) is an undersampled image with 94% of the image discarded (the size of the input image is 16 x 16). Its frequency domain representation in Fig. 4.10(j) exhibits the under sampled nature of the input by its spectrum repetitions within the image window. The major purpose of including this example is to infer the applicability of the solution technique for a different measurement function A and hence a different Q function. Therefore, the input samples are assumed to be in the frequency domain. Applications of Fourier synthesis in radio-astronomy belongs to this class and the measured inputs are of the form (without noise error terms)

$$D_k = \sum_{j=1}^{n} f_j \exp\left(-i\frac{2\pi jk}{n}\right) \text{ for } k = 1,\ldots\ldots m;\ m < n \qquad (4.10)$$

where D_k is the complex frequency sample and $i = \sqrt{-1}$.. The function A in this case is the complex exponential basis function with sine and cosine frequency components. Accordingly, in the procedure, the function Q is defined in terms of complex frequency values at known sample points. That is

$$Q(\hat{F}) = \sum_{k \in \mathbf{D}} |D_k - \hat{F}_K|^2 \qquad (4.10.1)$$

where $\mathbf{D}$ is the domain formed by the known frequency samples. $\hat{F}_k$ is the k^{th} frequency component of the estimate. The ill-posed nature of this problem class is due to incompleteness.

4.3.5. Results of the *ME* estimation

The *ME* relaxation algorithm was run on the five examples explained above and the

results upon convergence are shown in Fig.4.11(d)-(f) for the first category of inverse problems. The input images are also shown for comparison. The features of the eye image appear notably enhanced in Fig.4.11(d). The *ME* performance, more appropriately, the performance of the *ME* method in a relaxation approach, needs further investigations in this example. Discussions are postponed to the end of this chapter. Evidences for restoration come from the severely distorted cases shown in Figs. 4.11(e) and (f). Particularly in 4.11(f), it may be noted that the results have faithfully followed the 'test gaps' of the source checker board pattern.

In running the algorithm on the noisy image, no noise variance was specified. Fig. 4.11(g) shows the *ME* estimate. The justification for noise reduction comes from the residual test, already explained in Fig 4.7(b). More tests in the study of the noise performance are discussed in the following section. Comparison of Fig.4.8(a) with Fig. 4.11(h) shows a satisfactory reconstruction performance. All these results are evaluated quantitatively for their performance further in section 4.3.8.

4.3.6. Restoration in presence of noise-discussion

One of the disturbing factors with the result in Fig 4.11(g) is that it pertains to iteration 6, which as can be seen from the parameter variations in Fig. 4.4 (l), is not a converged result. Visually , the results upon and after convergence did not appear smooth. To reason out this behaviour quantitatively, the algorithm was run on the same input but with lower noise level (signal to noise ratio of 16.49 dB) shown in Fig. 4.12 (a). In this case, the results have shown a satisfactory performance upon and even after convergence. The result at iteration 10 is shown as an example in Fig 4.12(b). For both cases (9.17 and 16.49 cases), the variations of the standard deviation of the noise residuals are calculated and plotted as a function of iterations in Fig 4.13. In the 16.49 dB case, the residual standard deviation is close to unity upon convergence. The 9.17 dB case on the other hand, shows unit value at iteration 6, which corresponds to the visually pleasing result. Upon and

after convergence, the histogram shows a value close to 0.5 as implied by the narrowness of the residual histograms. This means that the agorithm, in the process the relaxation, attempts to fit the input data too closely after some stage But at iteration 6, the deconvolution performance would be unsatifactory. Hence the results did not appear smooth in the case of 9.17dB upon convergence. It follows therefore that in presence of noise terms, there is an optimum level of performance depending on the level of noise in the input, which can be found by the algorithm from the residual standard deviation calculations. It is however anticipated that the noise optimum problems can be resolved efficiently with suitable design techniques to control the convergence process for the noise terms independent of the data.

4.4. PERFORMANCE EVALUATION

The purpose of this section is to evaluate the usefulness of the *ME* method in applications related to inverse problems. Evaluations of this type are in general carried out by comparative studies with the standard techniques for a given application. Here, the deconvolution problem is considered and the performance of the *ME* method is studied along with two standard algebraic techniques of restoration namely, inverse and the constrained least squares (minimum mean square error) filtering techniques [28], [29], [30], [39]. The worst case images in Fig 4.8(b),(c) and 4.8(a) are chosen as inputs for the study.

4.4.1. The algebraic techniques of restoration

In inverse filtering approach, an image is restored in two steps, first, by finding the Fourier transform of the true image using the relation

$$\frac{D_k}{H_k} = \hat{F}_k \qquad (4.11)$$

In the discussions of this section, the upper case letters denote the discrete Fourier transforms of the corresponding spatial signals denoted by lower case characters. In standard restoration terminology, h is the point spread function (one of the measurement functions A in the *ME* method), k is a frequency variable that stands for (k1, k2) in two dimensions. The transfer function of the inverse filter in (4.11) is $1/H_k$. Performing the inverse fourier transform operation

$$\hat{f}_j = F^{-1}(\hat{F}_k) \qquad (4.11.1)$$

gives rise to an estimate of the true image. The symbol F^{-1} denotes an inverse Fourier transform operation. In the regions where the values of H_k are zeroes or small, the restored result becomes unreliably high. If both H_k and D_k happen to be small at high frequencies, then the image quality becomes severley distorted. In practice, the difficulty is overcome by truncation of H to its limited neighbourhood, eliminating the values close to 0. The performance of inverse filtering will be worse in presence of noise because the filter design does not include the underlying noise process. In presence of noise, one can use the Wiener filtering approach. Assuming additive noise, the transfer function H_r of the restoring filter in this approach is given by

$$H_r = \frac{H_k^*}{[|H_k|^2 + \phi_{nk}]} \qquad (4.12)$$

H_r can be used in the same way as in (4.11) and (4.11.1) to find the true image estimate. ϕ_{nk} is the noise power spectral density term. An improved form of (4.12) results from a stochastic process model for the ideal image field with known power-spectral density of ϕ_{fk} and a zero mean. The restoration filter in this case has been shown to be

$$H_{rk} = \frac{H_k^*}{\left[|H_k|^2 + \frac{\Phi_{nk}}{\Phi_{rk}} \right]} \qquad (4.13)$$

In practice, the image field power-spectral density will be unknown and there are many variations of (4.13) reported in the literature. One such technique is the constrained least-square filter developed by B.R. Hunt [39], [25]. It is a parametric restoration filter with a transfer function of the form

$$H_{rk} = \frac{H_k^*}{\left[|H_k|^2 + \rho |C_k|^2 \right]} \qquad (4.14)$$

ρ is a design constant and C_k is a design spectral variable constraint related to the human visual system.

4.4.2. Comparative study

The two techniques were applied to the worst case images along with the *ME* method. The results are shown in Fig. 4.14. As pointed out earlier, the fundamental difficulty with the chosen filter is that its transfer function has many zeroes in its pass band. The inverse filter is therefore not physically realizable and a large truncation approximation is necessary. Expectedly, the performance of the inverse filter is not satisfactory. The least squares filtering improves the situation but is still not satisfactory. Improvements may however be possible with carefully chosen design parameters.

The techniques being non-iterative, are computationally simple compared to the *ME* methods, but functionally have many drawbacks. In both techniques, the practical truncation limits and the design parameters are often found on a trial and error basis. As a result, the functional performances of these techniques vary significantly with varying inputs and blur conditions. Adjusting the parameters for the required performance becomes difficult and impractical. The term 'required performance' needs special attention. It refers to a subjective assessment, which as noted all along, is acceptable on the grounds of lack of

a quantitative measure for the quality of an image. But, it seems to have a more significant role for the techniques. The techniques rely upon the subjective evaluations and the understanding of 'image quality' for the possible design improvements. This implies that the lack of understanding of image quality has a direct impact on the progress of the techniques. A reasonable way to improve the filter design techniques is to improve the existing techniques based on some definite purpose, or to use ad-hoc techniques to study their effects on the final results. Unfortunately, in both cases, there lies ahead the problem of distinguishing the artifacts from the useful features.

The results obtained using the *ME* method are shown in Fig. 4.14(d). They are relatively free from ringing artifacts of inversion. More importantly, the method does not require any ad-hoc procedures for its functioning. Yet, its performance shows consistency with the different types of inputs considered for the study. The objectivity of the solution (explained in section 4.3.2) and the miminum interference of the artifacts contribute significantly to the visual evaluation of the results. The result obtained with the eye image in Fig. 4.11(d) indicates this possibility. Starting from the visual evaluation results of Fig. 4.11(d), significant developments become possible by using the *ME* method. It appears that the computational tediousness of the method is not a major issue compared to the magnitude of the problems of inversion. On the other hand, its impact on implementation is a matter of concern, but not a problem from the stand point of current technological developments. Studies show that the current technological developments are encouraging for the implementation of the *ME* method in general, particularly, for relaxation schemes. Along this line of reasoning, the relaxation approach proposed in this thesis work shows a promising direction for implementation. The implementation issues are covered in chapter 7.

4..4.3. Mean square error evaluation

In practice, many image processing applications make use of image fidelity

measures for performance evaluation. One of the simplest and standard measures of performance with discrete (sampled) monochrome images is the normalized mean square error measure [28]. The measure is given by

$$NMSE = \frac{\sum_{j=1}^{n} (f_j - \hat{f}_j)^2}{\sum_{j=1}^{n} f_j^2} \qquad (4.15)$$

where the f_j represents the true data and the $\hat{f}_j$ represents the estimated data. In cases where the information on true image is not available, f_j and $\hat{f}_j$ are replaced by the input data d_j and $\hat{d}_j$, with the latter calculated using $\hat{f}_j$. Some times, the measure is conveniently expressed as a signal to noise ratio in dB using the relation

$$MS(SNR) = -10 \log_{10}(NMSE) \qquad (4.10.2)$$

The higher the dB value, the better is the performance of a method in following the true signal. Based on (4.10.1), the measures are calculated for the input and the restored images in Fig. 4.11(a)-f and 4.11(a)-(d). In these calculations the wrap around error terms are eliminated since these error terms are inevitable in a convolution operation. Fig. 4.15 shows the performance of the *ME* method alone for the inputs 1-5. The inputs 2 and 3 and 5 have improved by 40%, 86.4% and 86% respectively by using the *ME* method. The increased error level with inputs 1 and 4 is as expected, because the results show certain characteristics that need visual evaluations and further investigations. The mean square error measure does not include these characteristics. Fig. 4.16 shows the results of comparative studies. The error level in the *ME* estimate for input-2 has reduced by 58% with respect to the constrained least squares approach. For input-3 it shows a reduction of 87%. The result with input-4 is as noted above, ambiguous because of the above mentioned discrepancies.

4.5. OBSERVATIONS FOR FURTHER INVESTIGATIONS

As noted earlier in section 4.3.5, the algorithm has restored the eye image in 4.11(d), with marked enhancement. The result pertains to iteration 9. The results at few other iterations, 2, 4 and 12 are shown in Fig. 4.17(a), (b) and (c) to show the sequences of changes in the process of restoration. The appearance of the iris at iteration-4 and the predominance of many features at iteration-12 gives rise to the possibility of the presence of an enhancement mechanism hidden in the procedure, which becomes operative in the relaxation process of restoration. This behaviour is also seen with the example in Fig. 4.12(b) that has minimal blur and noise. The fact that these examples have minimum levels of degradation leads to the formulation of a special test for the procedure. The simple pattern shown in Fig. 4.18 (a) was generated for this purpose. The basic idea behind the test was to investigate the performance of the algorithm with no degradations in the input data. Such a case does not fall directly under the class of inverse problems in (3.1). In order to run the algorithm on the undegraded input, the input samples are assumed in the frequency domain and the algorithm performs restoration based on (4.8) with $m = n$. The results of this test showed an unexpected behaviour. They are shown in Fig. 4.18. The bright and dark bands that appear (faint appearance) in the input image in Fig4.18(a) are due to our perception well known as Mach's bands. The predominant bright and dark bands appearing in Fig. 4.18 (b), on the other hand, are the results of the *ME* method using a relaxation algorithm with no degradations. This observation is the motivating factor for the next chapter.

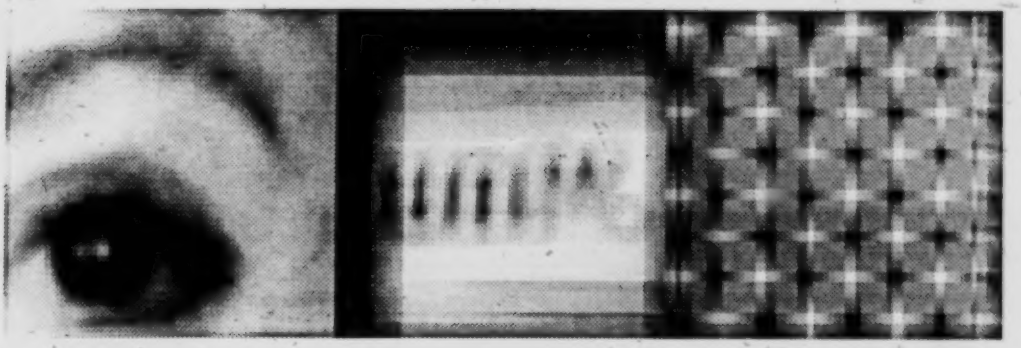
Algorithmic activity:	Input	Output
PHASE I: input		
reading:	$\underline{d}$, A	$\underline{d}$, A
PHASE 2: data processing		
n=1		
initialisation:	$\underline{f}^{(0)}$	$\underline{f}^{(0)}$
data space calculations:	$\underline{f}^{(n-1)}$, $\underline{d}$	Q, $\partial Q^{(n)}/\partial \underline{f}$
parameter space calculations:	Q, $\underline{f}^{(n-1)}$	$\lambda^{(n)}$, $\beta^{(n)}$, $q^{(n)}$
estimate calculations:	$\lambda^{(n)}$,$\beta^{(n)}$,$q^{(n)}$, $\underline{f}^{(n-1)}$	$\underline{f}^{(n)}$
convergence measurement:	$\underline{f}^{(n)}$, $\underline{f}^{(n-1)}$	$\gamma^{(n)}$
PHASE 3: termination check		

If $\gamma^{(n)} \leq \gamma_0$ then terminate and output $\underline{f}^{(n)}$

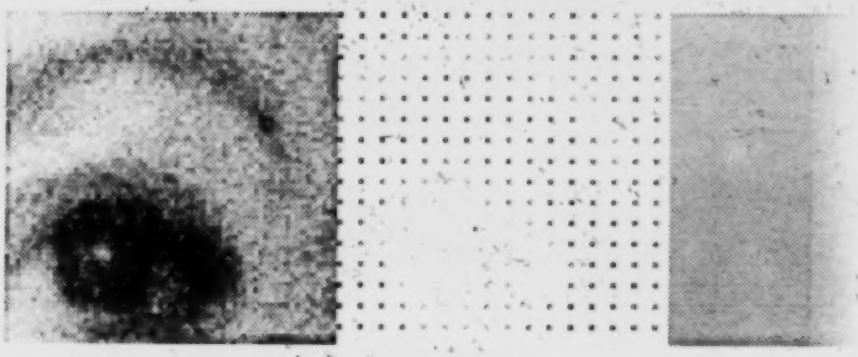
else $n = n+1$; $\underline{f}^{(n)} = \underline{f}^{(n-1)}$; repeat phase 2 and 3

2 - D arrays. (outlined) [e.g; $f = \{f_{j=(j1,j2)}; j1 = 1,\ldots\ldots, n1; j2 = 1,\ldots\ldots, n2\}$

Fig. 4.1 Algorithm for test studies



(a) (b) (c)



(d) (e) (f)

Fig. 4.2 Inputs to the algorithm

(a) input-1; (b) input-2; (c) input-3; (d) input-4; (e) input-5; (f) input-6.

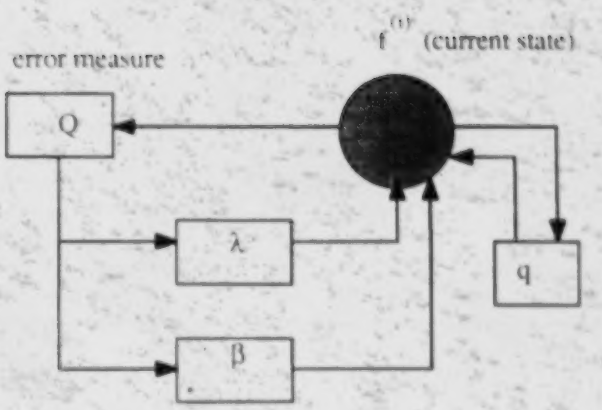


Fig. 4.3 Parameter control mechanism.

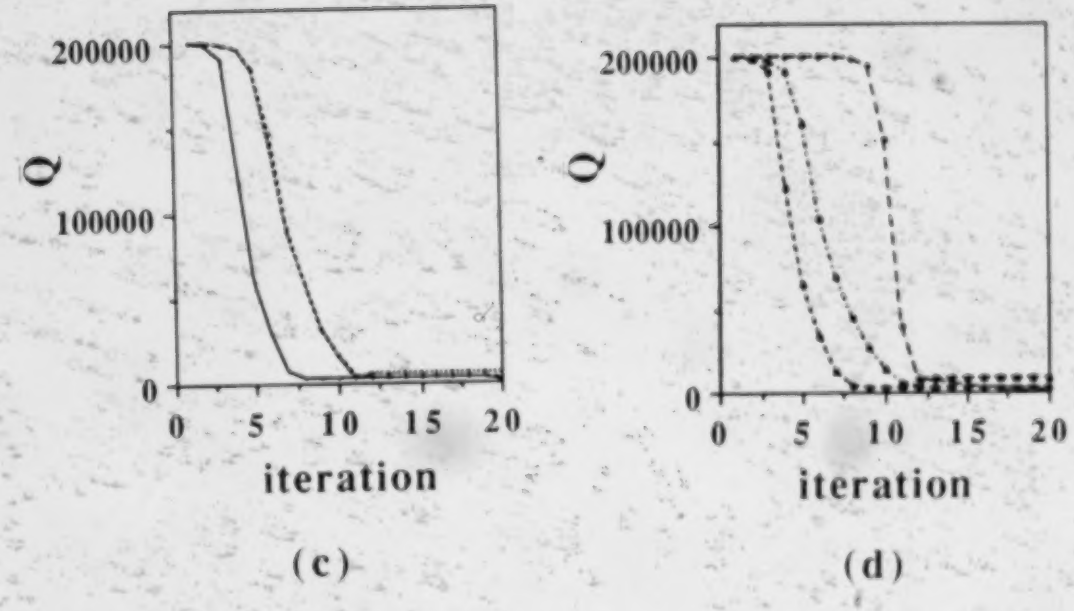


Fig. 4.4 Recorded parameter variations.
Q variations for (a) inputs 1 - 3; (b) inputs 4 - 6.

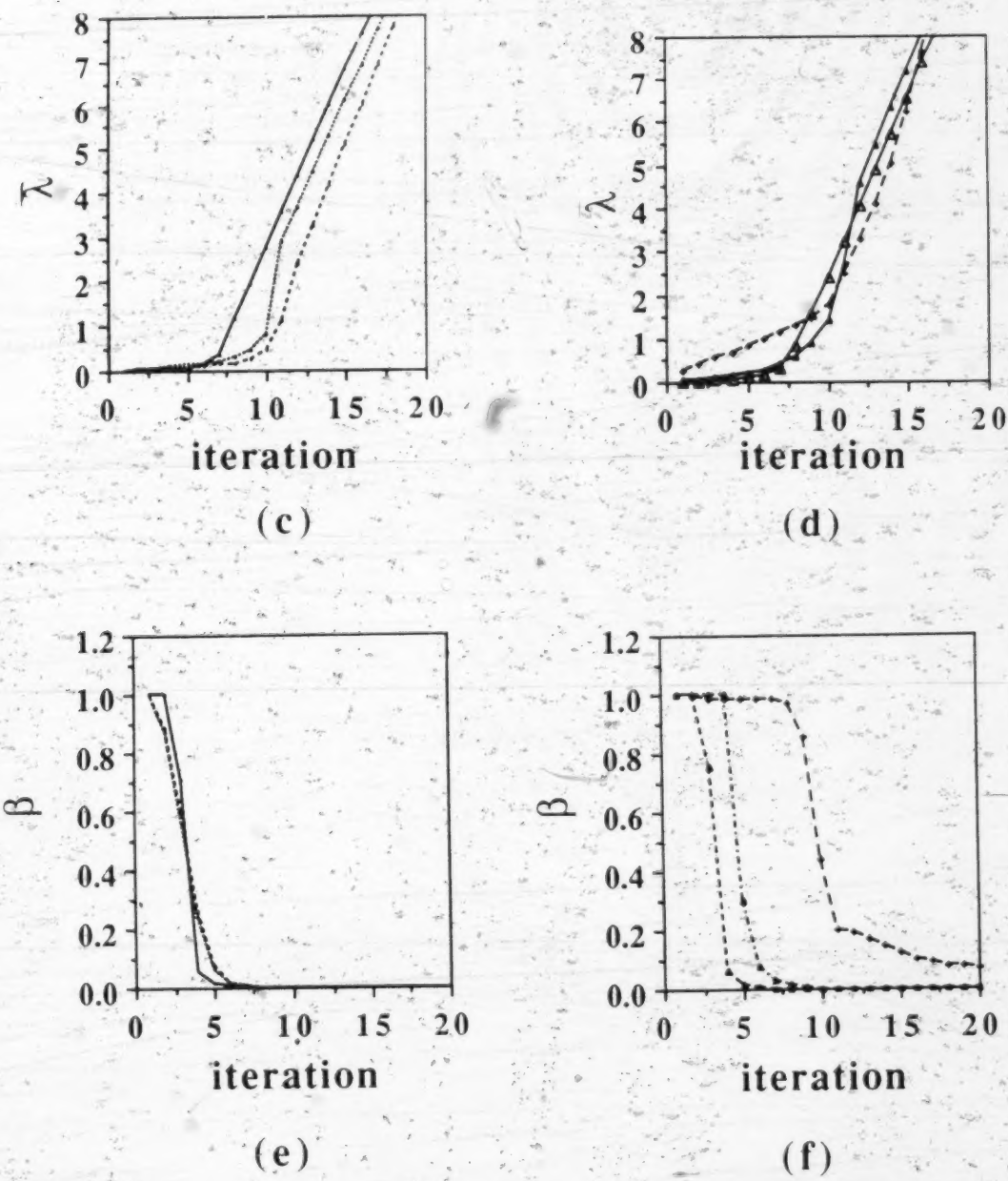
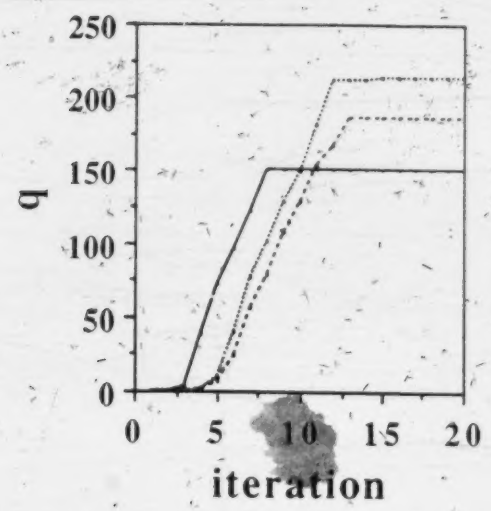
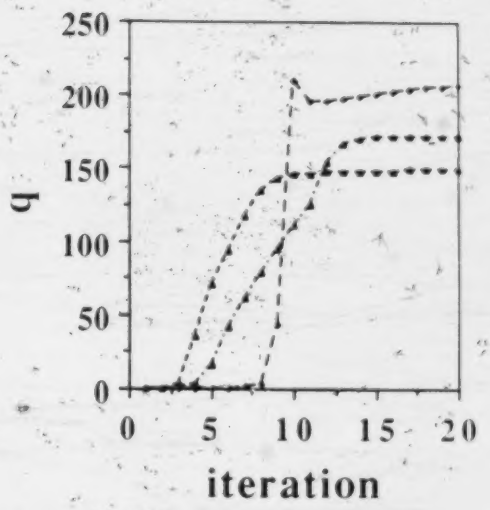


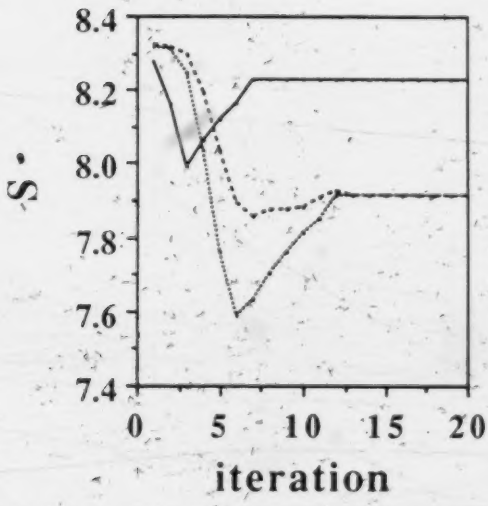
Fig. 4.4 continued. parameter variations.
Variations of $\bar{\lambda}$ for (c) inputs 1 - 3; (d) inputs 4 - 6.
Variations of β for (e) inputs 1 - 3; (f) inputs 4 - 6.



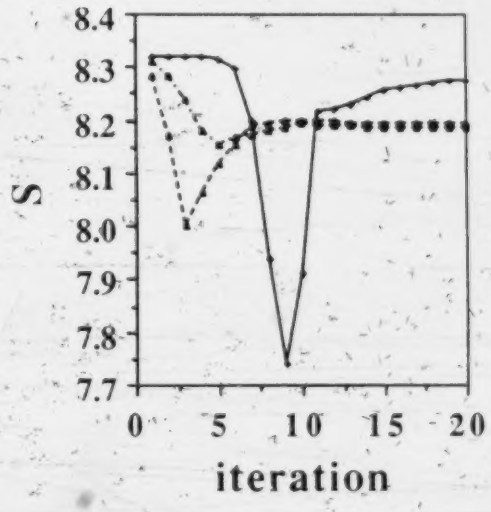
(g)



(h)



(i)



(j)

Fig. 4.4 continued. parameter variations.

Variations of q for (g) inputs 1 - 3; (h) inputs 4 - 6.

Variations of S for (i) inputs 1 - 3; (j) inputs 4 - 6.

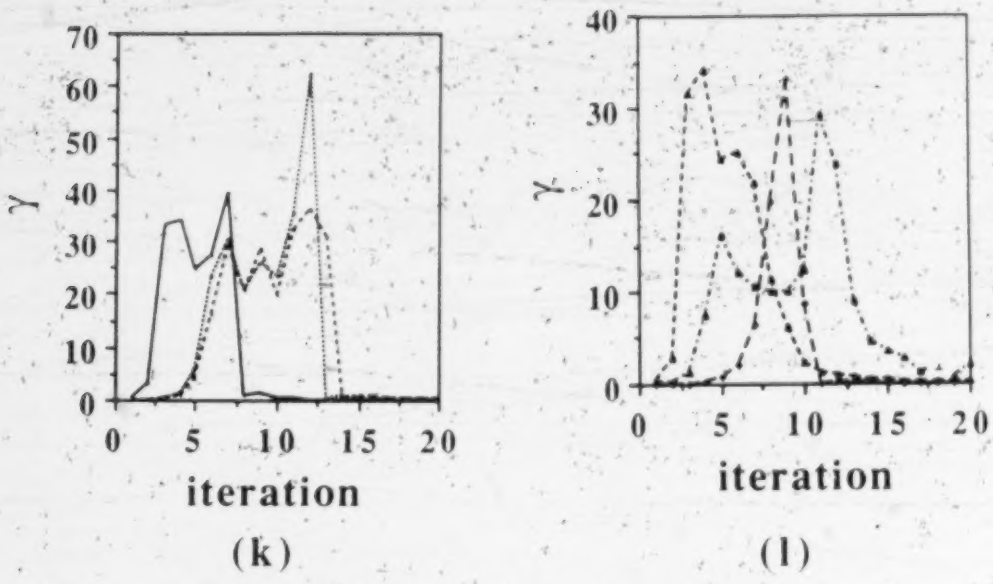


Fig. 4.4 continued. parameter variations.
Variations of γ for (k) inputs 1 - 3; (l) inputs 4 - 6.

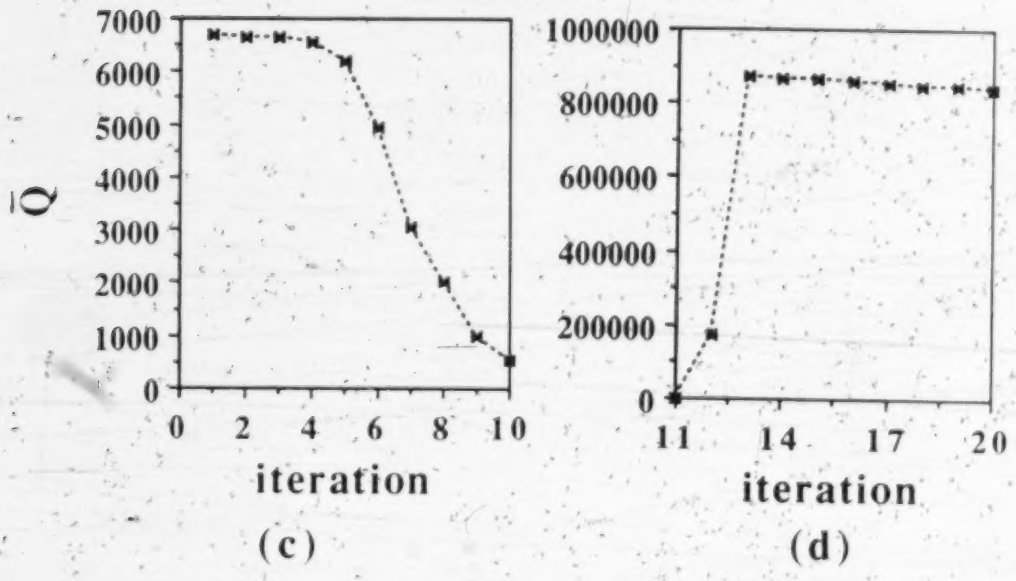


Fig. 4.5 Convergence test characteristics - removal of upper bound constraint.
Variations of $\bar{Q}$ in the (a) active phase; (b) termination phase.

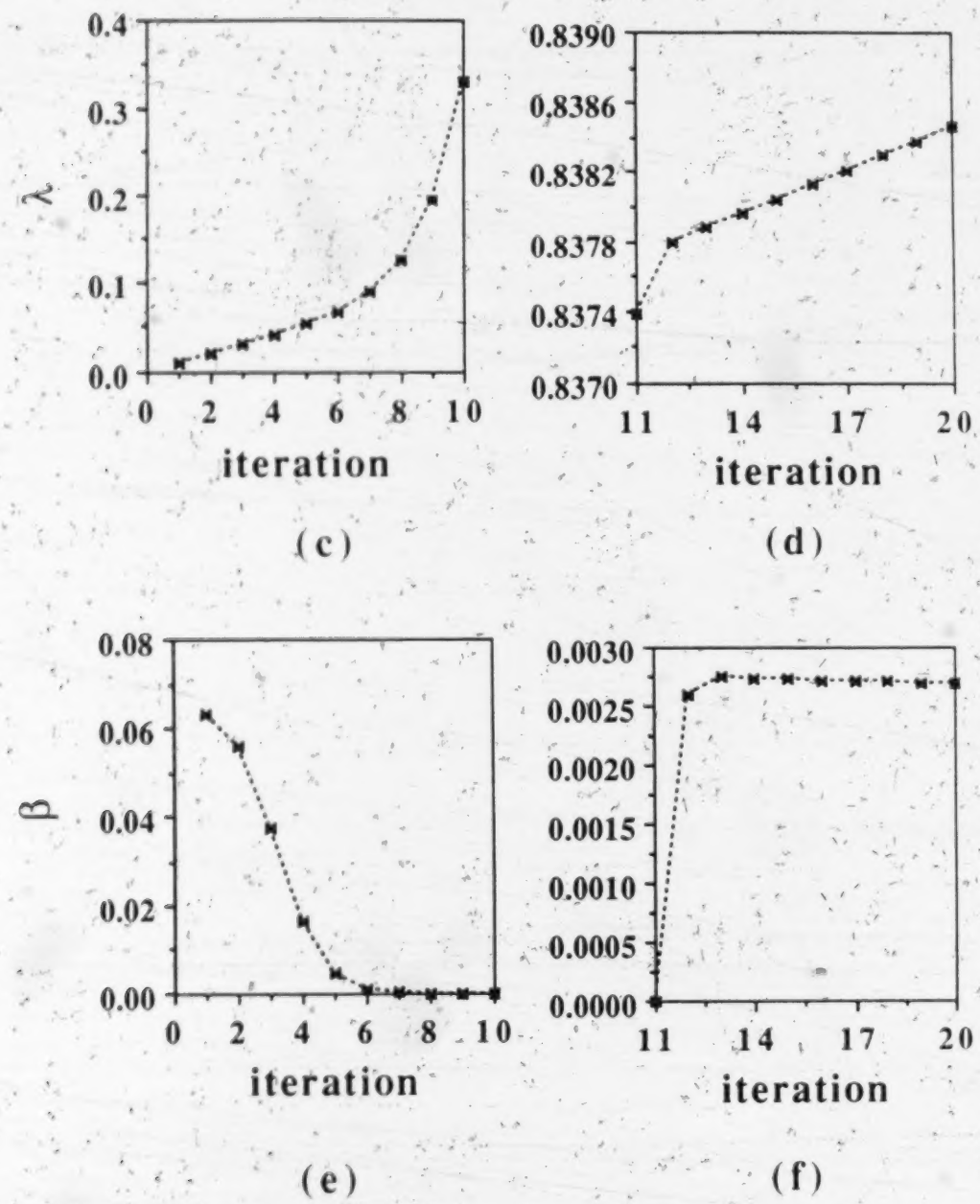


Fig. 4.5 Continued. Convergence test characteristics
Variations of $\bar{\lambda}$ in the (c) active phase; (d) termination phase.
Variations of β in the (e) active phase; (f) termination phase.

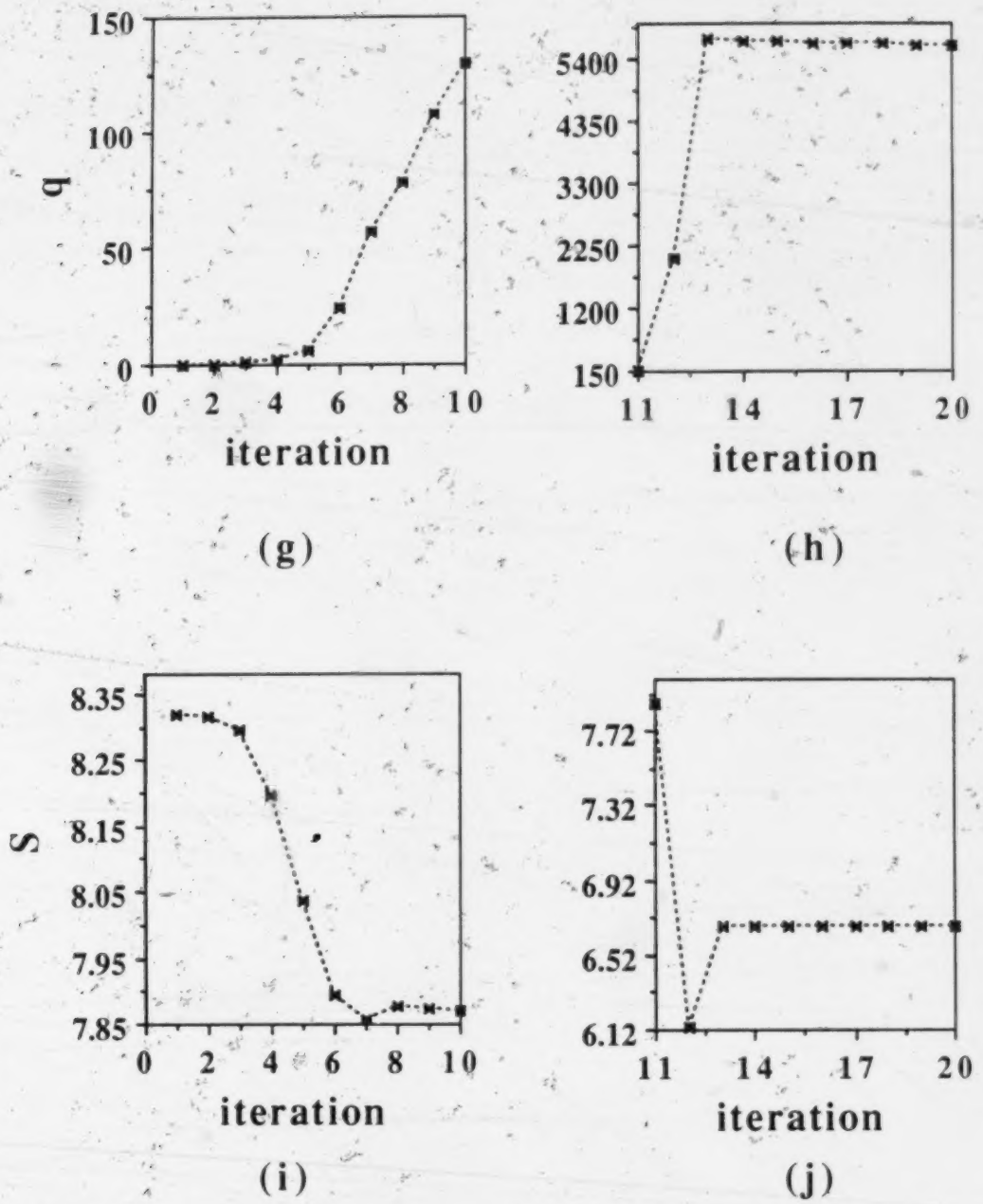


Fig. 4.5 Continued. Convergence test characteristics
Variations of q in the (g) active phase; (h) termination phase.
Variations of S in the (i) active phase; (j) termination phase.

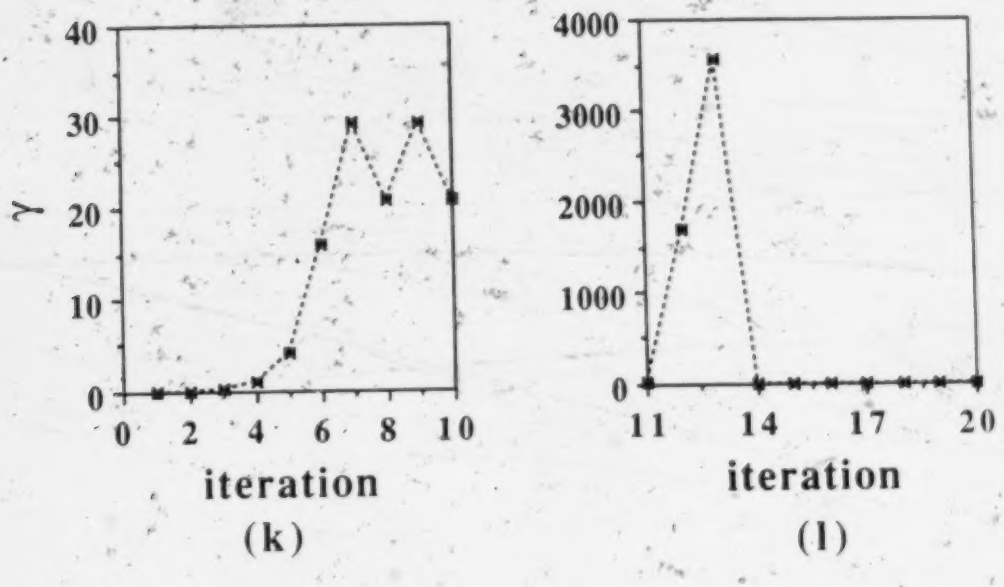


Fig. 4.5 Continued. Convergence test characteristics
Variations of γ in the (k) active phase; (l) termination phase.

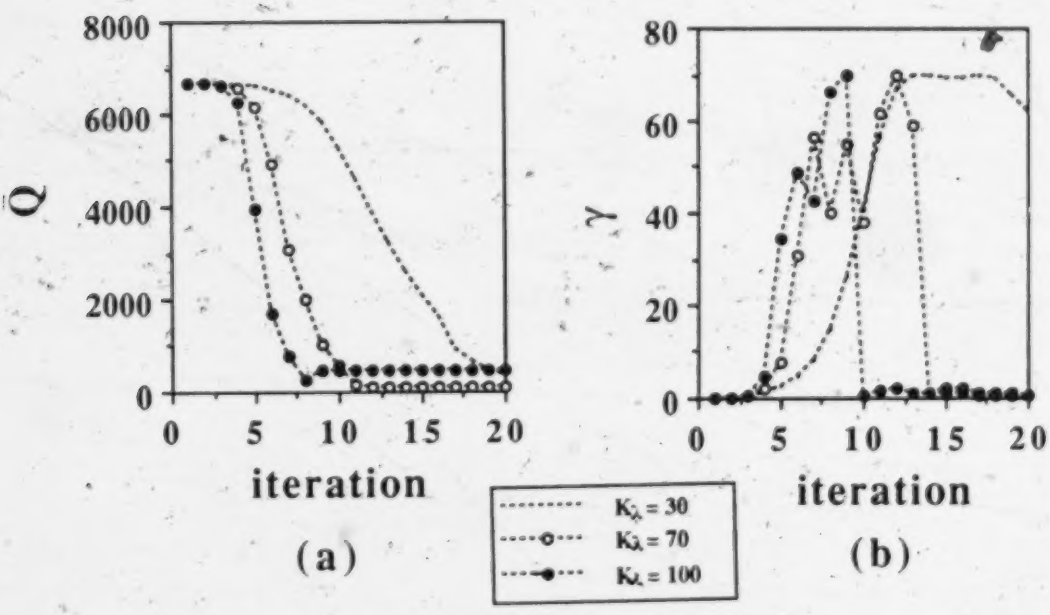


Fig. 4.6 K_λ Vs convergence speed.
Variations of (a) $\overline{Q}$; (b) γ.

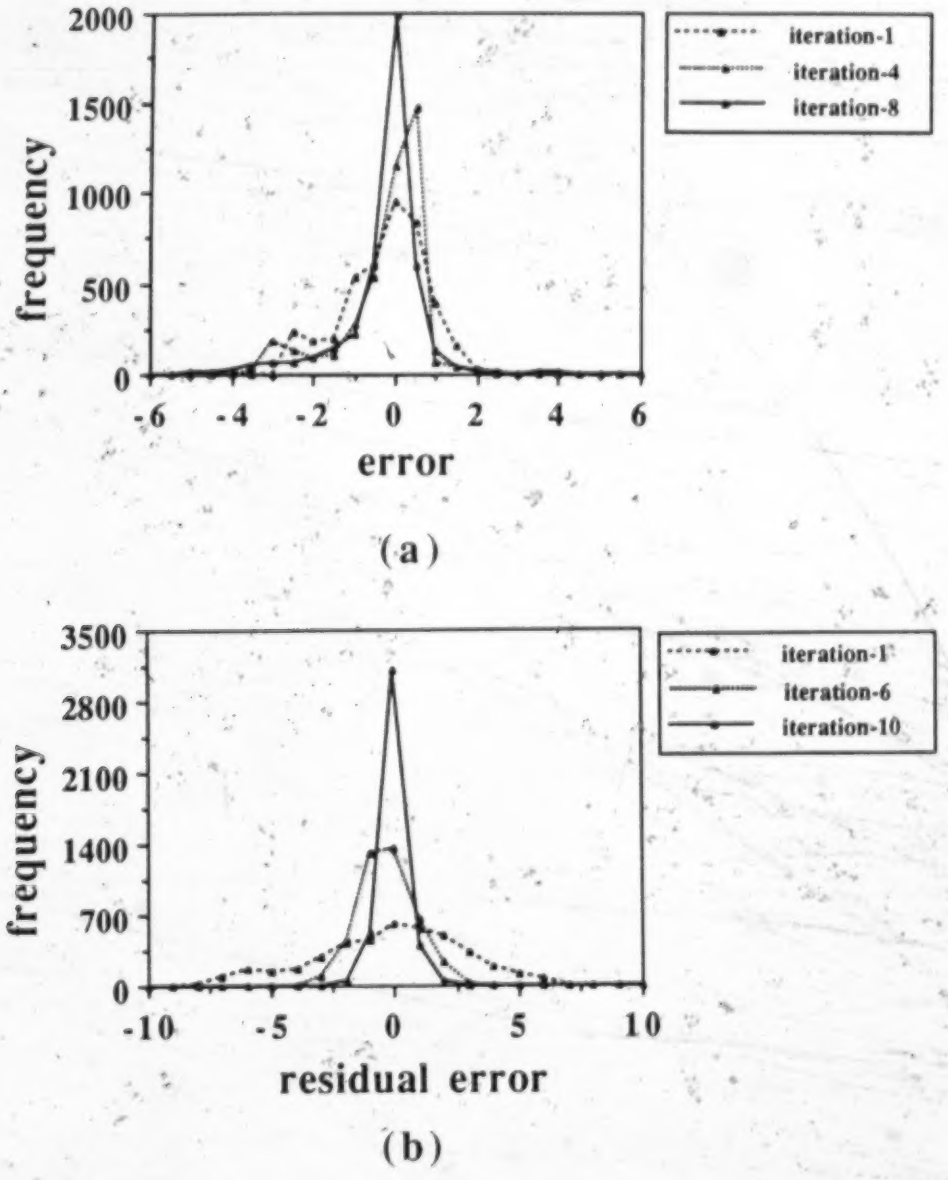
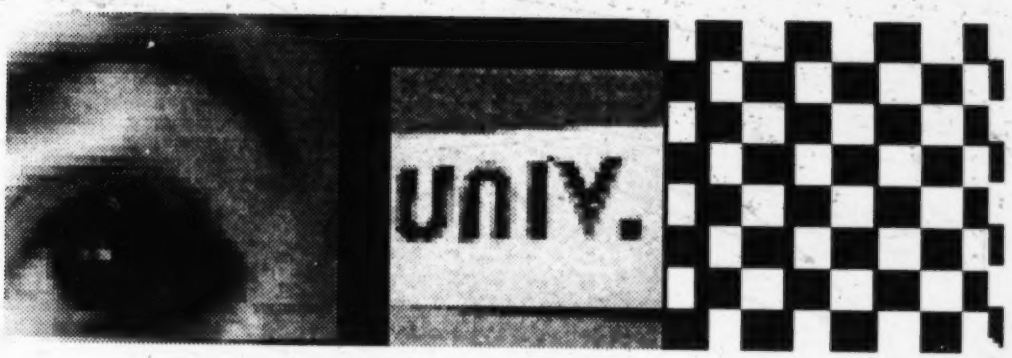
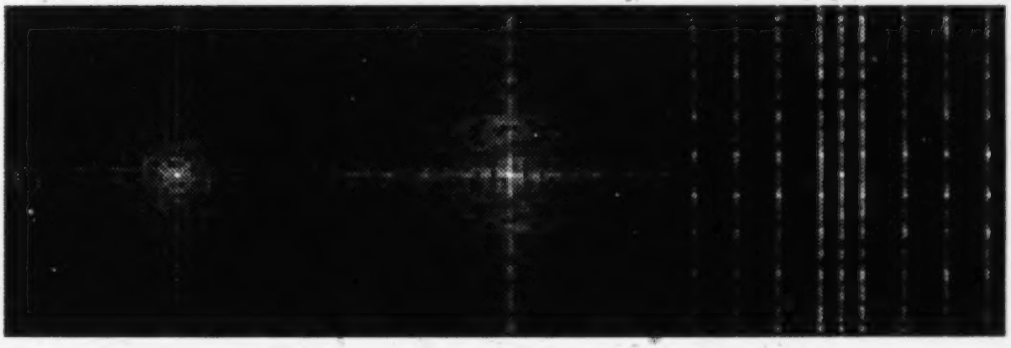


Fig. 4.7 Intermediate error histogram variations.

(a) in the absence of noise; (b) in presence of noise.



(a) (b) (c)

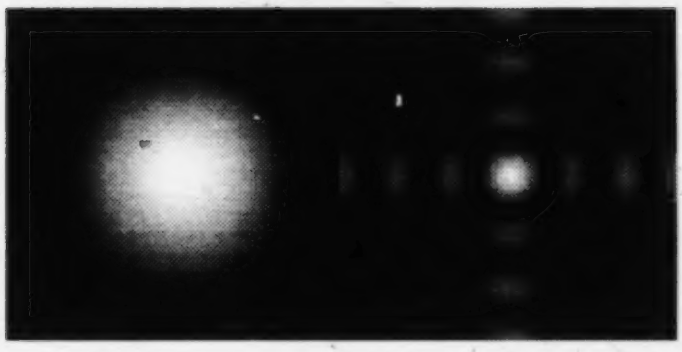


(d) (e) (f)

Fig. 4.8 Source images and their frequency spectra

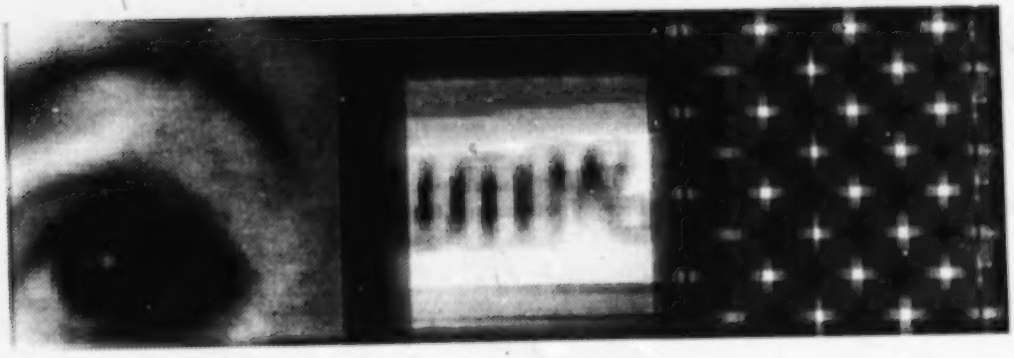
(a) image - 1; (b) image - 2; (c) image - 3.

Display of $\log(1+|F_k|)$: (d) image - 1; (e) image - 2; (f) image - 3.

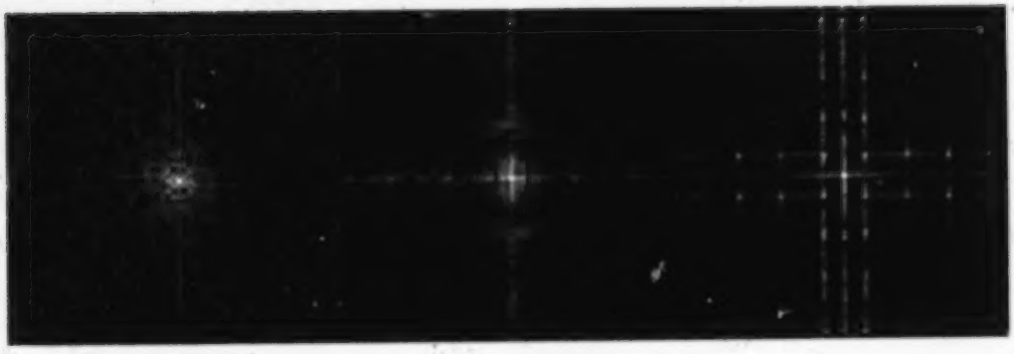


(a) (b)

Fig. 4.9 Transfer functions of the window filter
(a) 3 x 3 'normal' Gaussian filter; (b) a 7 x 7 'worst case' filter.

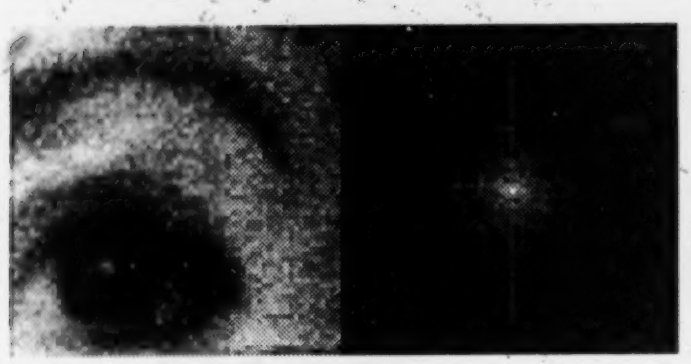


(a) (b) (c)

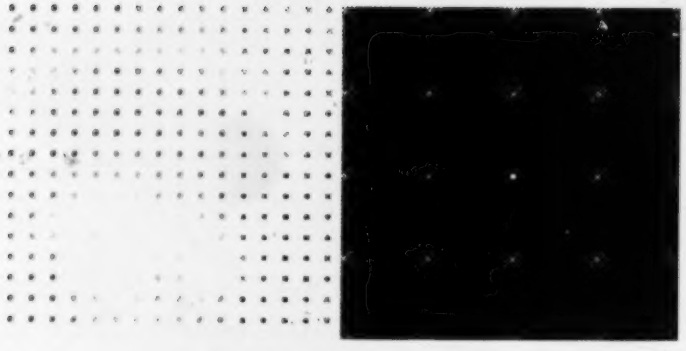


(d) (e) (f)

Fig. 4.10 Generated input images and their frequency spectra
input images: (a) input - 1; (b) input - 2; (c) input - 3.
Frequency spectra of: (d) input - 1; (e) input - 2; (f) input - 3.

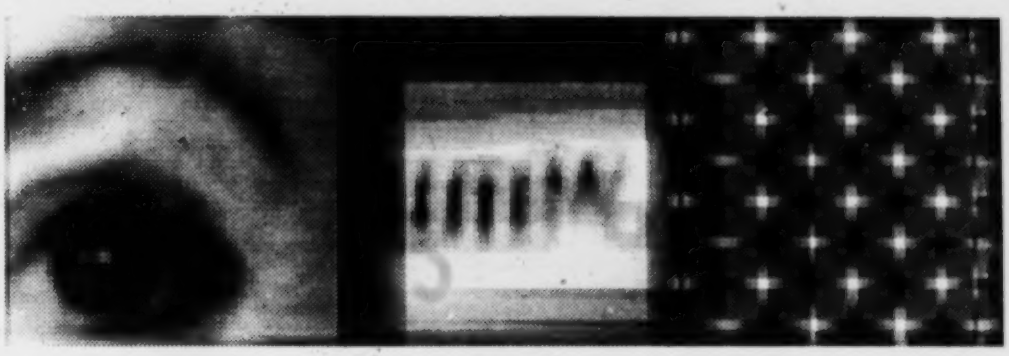


(g) (h)



(i) (j)

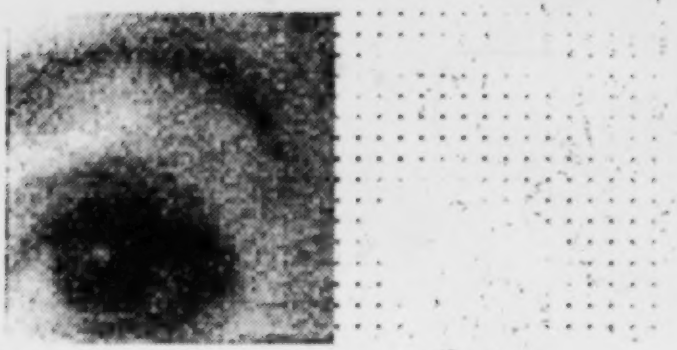
Fig. 4.10 Continued. Generated input images and their frequency spectra input images: (g) input - 4 (blurred and noisy; S/N ratio = 9.17dB); (i) input - 5 (under sampled, 94% of the data discarded); Frequency spectra of: (h) input - 4; (j) input - 5.



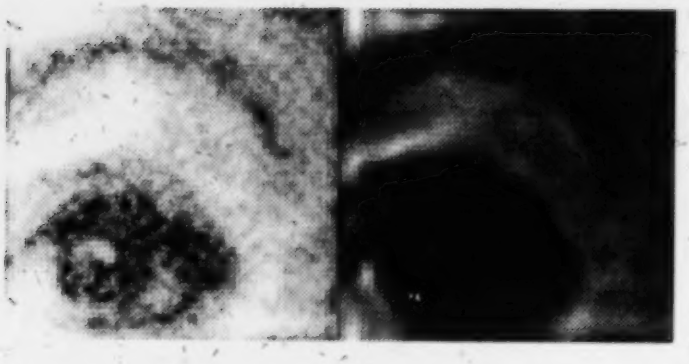
(a) (b) (c)

(d) (e) (f)

Fig. 4.11 The degraded inputs and the *ME* restored images
input images: (a) input - 1; (b) input - 2; (c) input - 3.
restored images of: (d) input - 1; (e) input - 2; (f) input - 3.



(g) (h)



(i) (j)

Fig. 4.11 Continued. The degraded inputs and the *ML* restored images
input images: (g) input #4; (h) input #5;
restored images of: (i) input #4; (j) input #5.

(a) (b)

Fig. 4.12 The *ME* performance for reduced noise level input
(a) input (S/N ratio = 16.49 dB) (b) *ME* restored result at iteration - 10

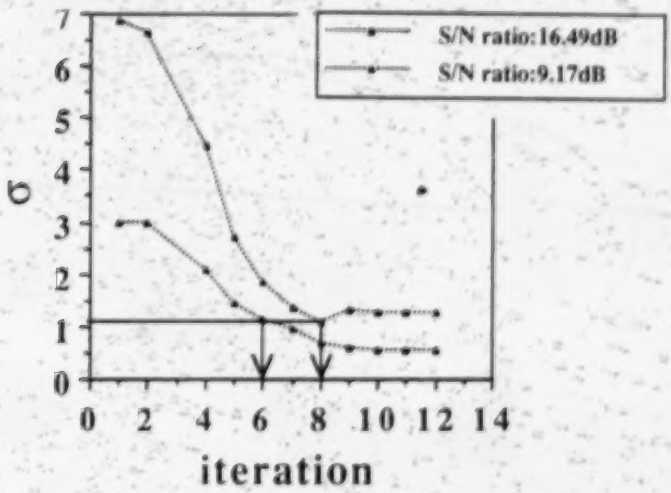


Fig. 4.13 Noise performance characteristics in the observed results.

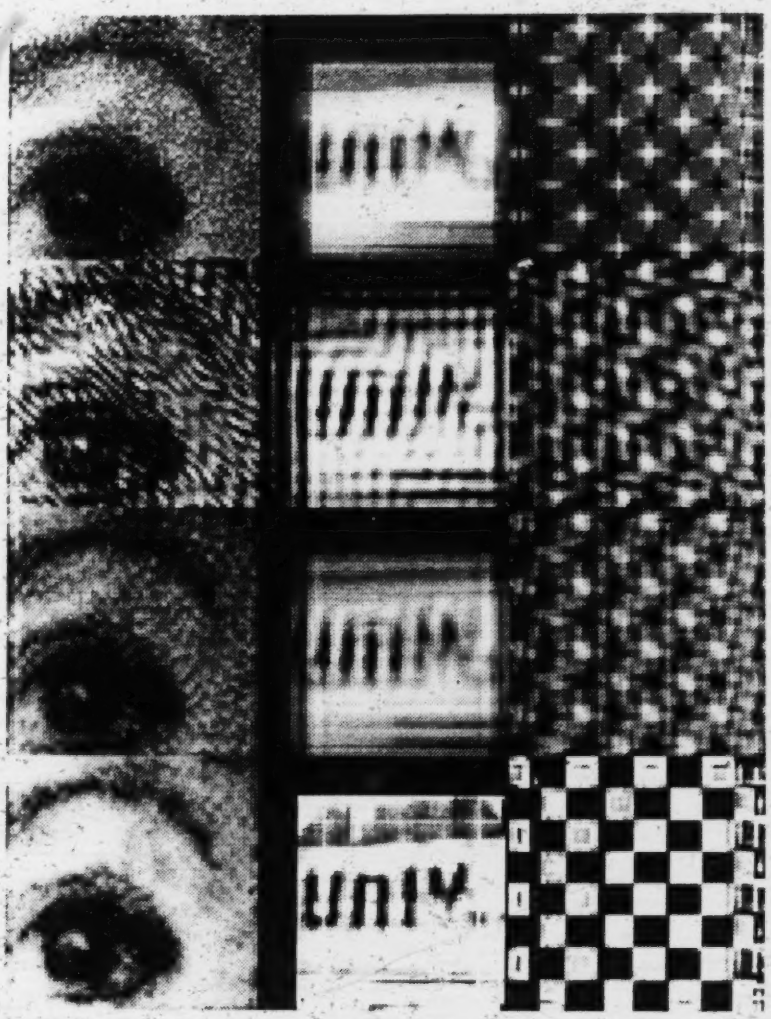


Fig. 4.14 The *ME* Vs algebraic restoration
(a) Degraded inputs, (b) Inverse filtering, (c) Constrained least squares filtering, (d) *ME* method

Fig. 4.15 Performance of the *ME* method in restoring inputs in Figs. 4-11

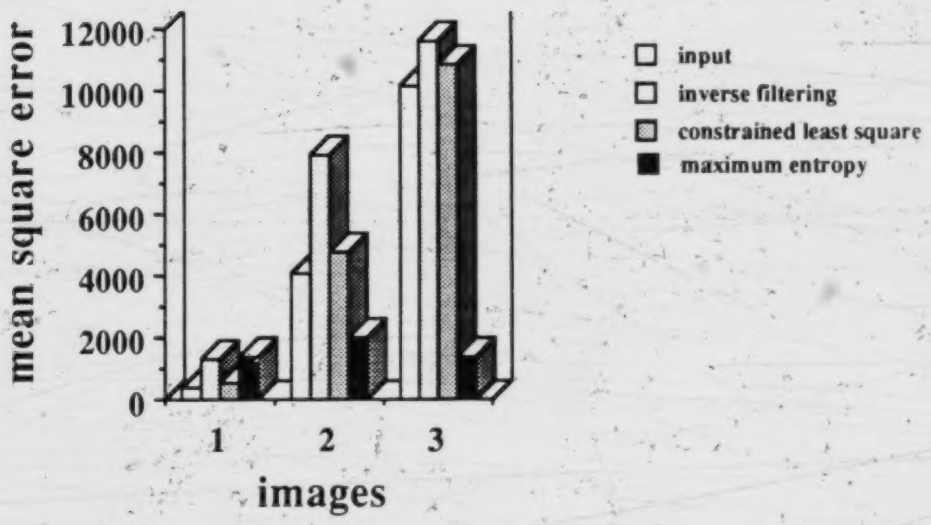


Fig. 4.16 Performance of the *ME* method in comparison with the standard methods

(a) (b) (c)

Fig. 4.17*ME* iterative improvements for the eye image in Fig. 4.11(d)
(a) iteration-2; (b) iteration-4; (c) iteration- 12.

(a) (b)

4.18 A test pattern and the observed *ME* estimate (a) input-6; (b) output at iteration - 10.

CHAPTER 5

MACH BAND BEHAVIOUR OBSERVATIONS - A SPECIAL TEST CASE STUDY

Test results of the simple pattern in chapter 4 are investigated in this chapter. The primary objective of this study is to identify the underlying mechanism (if any) that led to the formation of the unexpected additional bands in the test pattern. This is done in three phases: in the first phase, experiments are planned and carried out to obtain more evidence. Interestingly enough, results show striking similarities to some of the psycho-physical characteristics of the Mach band phenomenon in biological visual systems. In the second phase, the procedure is investigated analytically to identify the underlying enhancement mechanism. Using the results of this test study, attempts have been made to account for the various observations in the third phase.

5.1. TEST STUDIES AND DISCUSSIONS

The 2-D intensity gradient patterns similar to the one shown in Fig.4.18(a) are used for most of the test studies reported in this chapter. Following the terminology of biological models, they are some times referred to as the 'stimulus' patterns. Fig. 5.1(a) shows the test features of the pattern. It is a computer generated pattern of dimensions 64 x 64 with the gradient area of 10 x 64 pixels, with two edges, e1 and e2. The intensity slope of the pattern is the ratio between the illuminance intensity difference and the width of the gradient region. The uniformity of the intensities in one of the directions (here Y-direction) allows an elegant representation of the inputs as 1-D intensity profile distributions. The test studies, as discussed under section 4.5, involve running the algorithm shown in Fig. 4.1 with $\partial Q/\partial f$ suitably modified for the type of input. The result is a *ME* estimate of the input pattern, sometimes referred to as the 'response' pattern.

5.1.1. Test results of the *ME* estimation

The *ME* estimate of the pattern in Fig. 5.1(a) was already shown in Fig. 4.18(b). The pattern was observed in the absence of the restoration process in the procedure. The 1-D intensity profile of the estimate is shown in Fig. 5.2 along with the input for comparison. As expected, the response is uniform over the constant intensity regions. Near the edges, the response shows unexpected overshoots and an undershoot. The overshoots have appeared near the convex and concave edges and the undershoot near the concave edge alone. Further, they are not observed exactly at the edge points e1 and e2, but close to them.

The responses with varying input patterns were studied using a set of patterns similar to 5.1(a), but with different gradient slopes. The gradient areas of the patterns were 0, 5, 10, 20 and 25 pixels wide in the X-axis direction and 64 pixels in the Y-axis direction. Their profile plots are shown by the plots in Fig. 5.3(a) and (b). The *ME* estimate for each of these patterns is shown in Fig. 5.4(a) and (b). In another test, the slopes of the pattern were modified by changing the illuminance intensities. The responses are shown for two different gradient widths in Figs. 5.5(a) [the inputs are not shown in this case for clarity]. All the test responses show overshoots and undershoots near the edges. More importantly, the widths and amplitudes of the overshoot and the undershoot responses have undergone a regular modulation depending on the gradient slopes. The characteristics of the overshoots and undershoots will be discussed shortly.

In the next test, the sensitivity of the estimate response was tested with the pattern shown in Fig. 5.6(a). It has 4 adjacent edges as close as possible to each other near the overshoot edge. The *ME* estimate of this pattern is shown in Fig. 5.6(b). The response shows evidences for the presence of all the adjacent edges. Their correpondences with the concavity or the convexity are not interpretable because of the shifts in the undershoots and overshoot responses. Nonetheless, the test results have given rise to an important clue.

5.1.2. Mach bands

The *ME* algorithm response to the selected test patterns are found similar to the Mach band behaviour in biological visual systems. The presence of enhancement mechanism in biological visual systems was first established by Ernst Mach (1838-1916) through the discovery of the Mach band phenomenon. The phenomenon shows the response of the biological visual system around the boundary regions of different intensities [29], [80], [81]. Whenever an edge pattern consisting of a black and white region (stimulus pattern) is presented to the eyes, two additional, unexpected thin light and dark strips will be perceived near the edges (this can be checked visually in the figure Fig. 4.18(a)). These unexpected additional bands are referred to as Mach bands. Using psycho-physical experiments, the characteristics of these bands were studied, the reports of which can be found in the literature [80]. The following section gives an account of the reported experimental evidences relevant to the test studies explained above. The information for the following section has been obtained from the book by Ratliff [80], which contains the translated version of the original works of Mach.

5.1.3. Mach band characteristics in biological visual systems

The characteristic of a typical Mach band as reported in the literature is shown as a schematic diagram in Fig 5.7 (a). The characteristic pertains to the stimulus pattern of type shown in Fig. 5.1(a). The amplitude and the widths of the bands have been found to vary with the slopes of the gradient in the stimulus pattern. Specifically, it has been reported that as the slope decreases, the width of the bright band becomes wider and disappears altogether for gradual slopes. The widths of the dark bands on the other hand, have been observed to remain constant over a range of slopes. The reports on the dark bands do not appear conclusive, perhaps because of the limited measurement capabilities of the equipment available during Mach's time. Also, the position of the bright band with respect

to the flexion point has been reported to be subjective, varying from observer to observer. The bands were found either symmetric around the point of flexion or into the bright field. Results however appear conclusive on the behaviour of the bright bands. More importantly, the widths of the bright bands are narrower compared to the dark bands. With respect to their positions, the two bands are strongly asymmetrical. The reported characteristics are shown in their schematic form in Fig. 5.8(b).

5.1.4. A comparison

Visual comparison of Figs. 5.4 with 5.7 and 5.8 shows strong similarities with respect to the asymmetry of the bands and the behaviour of the bright bands for the varying widths of the stimulus pattern. The bands, in one dimension, refer to the overshoots and undershoots of their intensity profile. The shift in the position of the bright bands is another notable feature. Comparisons are not attempted on the dark band characteristics, since the report evidences are understood to be inconclusive. However, the similarities of the *ME* algorithm response with the Mach bands near the bright band are supportive of the presence of a hidden enhancement mechanism in the relaxation procedure of the *ME* method. It is also quite possible that the mechanism is interpretable in terms of the models of biological visual enhancements reported in the literature.

5.1.5. Need for suitable parameter settings

Encouraged by the initial results with simple pattern, examples of complex perceptual test patterns typically used in Mach band studies [80] were considered for the test studies. Multi-gradient steps, staircase, simultaneous contrast test patterns are some of the kinds that were tried and the *ME* algorithm results were *not* found favourable. Following the discrepancey, the inital tests were rerun with modified values of β and λ. With the modified parameter values, the overshoot amplitudes and widths varied significantly and the results could not be interpreted. A sample response with varying β is

shown in Fig. 5.9(a). Infact, the response in Fig. 5.5(a) itself is the result of using a specific value for β, determined on a trial and error basis. These test conditions, combined with the fact that the responses do not have any external functional influence (absence of restoration) suggest that the parameter settings of the relaxation process have a major bearing in the Mach band functional performance of the *ME* algorithm. In this sense, the simple pattern case is only a specfic test case. The lack of a suitable parameter setting is one possible reason for the failure with the complex patterns. A possible strong reason comes from the *ME* method itself. The estimation approach for the above test conditions make use of the single constraint *ME* method. In this approach, the parameters λ and β are global to the stimulus pattern. Perhaps, the global parameter setting applies well to uniform stimulus patterns. If this argument is true, then the standard *ME* method with $\mathbf{m}$ constraint parameters in a relaxation approach is expected to yield favourable responses with complex patterns.

5.1.6. The behaviour conditions of the algorithm for Mach bands

One of the useful pieces of information for the study of Mach bands comes from *ME* response variations in the process of relaxation. Fig. 5.10(c) shows the sequences for an arbitrarily chosen pattern of slope 100/10 . It shows that the bands with typical Mach band features appear only at iteration 10. After the 10^{th} iteration, the dark bands begin to disappear, while the bright bands still persist. This observation turns our attention to the recorded behaviour characteristics of the procedure discussed in chapter 4 to find the 'uniqueness' of the iteration 10. The input-6 variations in Fig.4.4 provides the required information. The variations of entropy shows that between iteration 9 and 10, the phase when Mach bands are observed, the entropy measure has just crossed its active phase and entered the termination phase. Conclusions are however not obtainable from this result because of lack of enough evidences.

5.1.7. Test study inferences-summary

The observations reported in chapter 4 along with the test results of this chapter are summarized below before proceeding with the analytical studies. Input images with minimal blurring conditions, when subject to the *ME* restoration process under a parameter relaxation scheme, gave rise to estimates with markedly enhanced features. The degree of enhancement was observed to increase with increasing iterations. The enhancement effects were also observed in the *absence* of a restoration process, under undegraded input conditions, using the *ME* parameter relaxation algorithm. In the latter case, a specific input pattern showed similarities with the psycho-physical characteristics of Mach bands in biological visual systems. The similarity was marked by the asymmetry and the amplitude as well as the width modulation of the bright bands with varying slopes of the input pattern. Also, a difference was noted in the functional performance of the *ME* algorithm between in the presence and absence of restoration. Unlike the presence of restoration case, the enhancement effects diminished with increasing iterations in the absence of restoration case and the bands appeared at a specific stage, when the entropy variations were at the crossing stage of the active and the termination phases. All these observations pertained to a simple test pattern. With more complex patterns, the anticipated enhancement effects were not observed. This failure was attributed to the globalized nature of the parameter settings in the relaxation process, resulting from the use of the single constraint *ME* method. In general, the parameter settings of the procedure were found to have a significant functional influence on the responses.

5.2. ANALYSIS - STAGE I

The test conditions under which the enhancement effects were observed fall into three different classes. In the test studies with the simple pattern, no degradation effects or filtering were present and the parameter settings alone were found responsible for the enhancement effects. The restoration results of chapter 4 on the other hand, had the

restoration mechanism as well as the parameter settings. For analysis considerations, we can introduce an intermediate test class of the simple restoration without any parameter settings.

5.2.1. Test class-1

Let us consider the true (unaveraged) *ME* estimate at any iteration i, obtained as a result of applying the parameter relaxation algorithm. Rewriting (3.50.1)

$$\hat{f}_j^{(i)} = q^{(i)} \exp\left\{ \bar{\lambda}^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \right\} \tag{5.1}$$

The function $\bar{Q}$ corresponds to the undegraded input for the test class and is given by

$$\bar{Q}^{(i)} = \frac{1}{2n\,i} \sum_{j=1}^{n} \sum_{t=0}^{i} \left(\bar{\hat{f}}_j^{(t)} - f_j \right)^2 \tag{5.2}$$

$\bar{\hat{f}}_j^{(t)}$ is the intermediate estimate of iteration t. The actual calculation of $\bar{Q}$ in the procedure takes place in the transform domain and here it is written in its equivalent spatial form for analytical convenience. From (5.2)

$$\frac{\partial \bar{Q}^{(i)}}{\partial f_j} = \frac{\sum_{t=0}^{i} \left(f_j - \bar{\hat{f}}_{jt} \right)}{in} \tag{5.3}$$

$\bar{\hat{f}}_j^{(t)}$ has been replaced by $\bar{\hat{f}}_{jt}$ for notational convenience. $\bar{\lambda}$ assumes its usual form following the procedure requirements and is given by (3.48). For present discussions, let us denote it as a function ϕ_i . That is

$$\bar{\lambda}^{(i)} = \sum_{t=0}^{i} \frac{K_\lambda}{\bar{Q}^{(t)}} = \phi_i \text{ for } j = 1, \ldots\ldots, n. \tag{5.4}$$

with $\lambda^{(0)} = 0$. The index j is introduced to signify the fact that λ is global and is constant for all the pixels for a given t. Substituting (5.4) and (5.3) in (5.2) and replacing j by p (p

stands for a particular element) we get

$$\hat{\hat{f}}_p^{(i)} = q^{(i)} \exp\left\{ \frac{\phi_{pi}}{in} \sum_{t=0}^{i} \left(f_p - \hat{\hat{f}}_{pt} \right) \right\} \tag{5.5}$$

Alternatively

$$\hat{\hat{f}}_p^{(i)} = q^{(i)} \frac{\exp\left\{ \frac{\phi_{pi}}{in} \sum_{t=0}^{i} f_p \right\}}{\exp\left\{ \frac{\phi_{pi}}{in} \sum_{t=0}^{i} \hat{\hat{f}}_{pt} \right\}} \tag{5.6}$$

(5.6) can be expressed in a more general form as

$$\hat{\hat{f}}_p^{(i)} = q^{(i)} \frac{\exp\left\{ \sum_{t=0}^{i} \phi_{pt} f_p \right\}}{\exp\left\{ \sum_{t=0}^{i} \phi_{pt} \hat{\hat{f}}_{pt} \right\}} \tag{5.7}$$

where

$$\phi_{pt} = \frac{\phi_{pi}}{in} \quad \text{for } t = 0, \cdots\cdots, i. \tag{5.7.1}$$

(5.7.1) is the characteristic function for the test class - 1.

5.2.2. Test class-2

In this class, the true *ME* estimate is the result of restoration without the relaxation control parameter β. Also, the analysis does not assume any specific form for λ. This class is only of analytical interest because, in practice, test results are not feasible without control mechanisms. $\partial Q/\partial f$ pertaining to the class can be rewritten from (3.16) as

$$\frac{\partial Q}{\partial f_j} = \sum_{k=1}^{n} \left(\sum_{r=1}^{n} \hat{f}_r h_{jr} - d_j \right) h_{jk}. \tag{5.8}$$

The measurement function A has been replaced by h (point spread function). Substituting (5.8) in equation (3.13), the *ME* estimate for the class is:

$$\hat{f}_j = q \exp\left\{ -\frac{\lambda}{\sigma^2} \sum_{k=1}^{n} \left(\sum_{r=1}^{n} \dot{f}_r h_{jr} - d_j \right) h_{jk} \right\} \qquad (5.9)$$

The presence of the exponential allows (5.9) to be written in the form

$$\hat{f}_j = q \frac{\exp\left\{ \frac{\lambda}{\sigma^2} \sum_{k=1}^{n} d_j h_{jk} \right\}}{\exp\left\{ \frac{\lambda}{\sigma^2} \sum_{k=1}^{n} \dot{f}_k h_{jk} h_{jk} \right\}} \qquad (5.10)$$

Replacing j by p, (5.10) can be expressed as

$$\hat{f}_p = q \frac{\exp\left\{ \frac{\lambda}{\sigma^2} \sum_{k=1}^{n} d_p h_{pk} \right\}}{\exp\left\{ \frac{\lambda}{\sigma^2} \sum_{k=1}^{n} \hat{d}_k h_{pk} \right\}} \qquad (5.11)$$

where

$$\hat{d}_k = \dot{f}_k h_{pk} \qquad (5.11.1)$$

Expressing (5.11) in a simple form using ϕ

$$\hat{f}_p = q \frac{\exp\left\{ \sum_{k=1}^{n} d_p \phi_{pk} \right\}}{\exp\left\{ \sum_{k=1}^{n} \hat{d}_k \phi_{pk} \right\}} \qquad (5.12)$$

where

$$\phi_{pk} = \frac{h_{pk} \lambda}{\sigma^2} \qquad (5.12.1)$$

(5.12.1) is the characteristic function for test class - 2.

5.2.3. Test class-3

Let us now consider the practical form of the *ME* estimate in the procedure used in obtaining the restoration results in chapter 4. Referring back to (3.68), the unaveraged *ME* estimate at iteration i is given by

$$\hat{f}_j^{(i)} = q^{(i)} \exp\left\{ -\bar{\lambda}^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \right\} \tag{5.13}$$

Q at iteration i for restoration is given by (3.39) and can be written in terms of the present notations as

$$\bar{Q}_i = \frac{1}{2n} \sum_{j=1}^{n} \left[\frac{1}{i} \sum_{t=0}^{i} \left(\sum_{k=1}^{n} \hat{f}_{kt} h_{jk} - d_j \right) \right]^2 \tag{5.14}$$

$\partial\bar{Q}/\partial$fj can be written in its simple form as

$$\left(\frac{\partial \bar{Q}}{\partial f_j} \right)_i = \frac{1}{in} \sum_{t=0}^{i} \sum_{k=1}^{n} \left(\bar{\hat{d}}_{kt} - d_j \right) h_{jk} \tag{5.15}$$

where

$$\bar{\hat{d}}_{kt} = \hat{f}_{kt} h_{jk} \tag{5.15.1}$$

Substituting for $\bar{\lambda}$ from (5.4) and $\partial\bar{Q}/\partial$fj from (5.15), the true *ME* estimate in (5.13) becomes

$$\hat{\bar{f}}_{pi} = q_i \frac{\exp\left\{ \sum_{t=0}^{i} \sum_{k=1}^{n} d_p h_{pk} \Phi_{pt} \right\}}{\exp\left\{ \sum_{t=0}^{i} \sum_{k=1}^{n} \bar{\hat{d}}_{kt} h_{pk} \Phi_{pt} \right\}} \tag{5.16}$$

where Φ_{pt} is given by (5.7.1). As in other test classes, j has been changed to p in (5.16). Equivalently

$$\hat{\bar{f}}_{pt} = q_i \frac{\exp\left\{ \sum_{t=0}^{i} \sum_{k=1}^{n} d_p \Phi_{pkt} \right\}}{\exp\left\{ \sum_{t=0}^{i} \sum_{k=1}^{n} \bar{\hat{d}}_{kt} \Phi_{pkt} \right\}} \tag{5.17}$$

where

$$\Phi_{pkt} = h_{pk} \Phi_{pt} \tag{5.17.1}$$

Comparing (5.7), (5.12) and (5.17), we see that although the test classes differ, the

true *ME* estimates of the test classes reduce to a common form. Within the form, the classes are characterized by the ϕ terms and the type of variables involved. The type of functional operations however remain the same.

5.3. ANALYSIS - STAGE II

The test classes can now be expressed using a more general form:

$$\tilde{\tilde{s}}_p = q_p \frac{\exp\left\{\sum_{i=c_0}^{a} \Phi_{pi} s_p\right\}}{\exp\left\{\sum_{i=c_0}^{c_T} \Phi_{pi} \tilde{\tilde{s}}_i\right\}} \qquad (5.18)$$

s denotes the signal and Φ stands for an arbitrary function that convolutes with s in determining the reponse at point p. The signal is assumed to be finite wiph limits c_0 and c_L. The exponential argument ratio term in (5.18) can be expressed as

$$\tilde{\tilde{R}}_p = \exp\left\{\sum_{i=c_0}^{c_T} \Phi_{pi} s_p - \sum_{i=c_0}^{c_T} \Phi_{pi} \tilde{\tilde{s}}_i\right\} \qquad (5.19)$$

(5.19) can be written as an inequality relation since $R_{pi} \gtrless 1$. Equivalently

$$\tilde{\tilde{\mathcal{A}}}_p = \left\{\sum_{i=c_0}^{c_T} \Phi_{pi} s_p - \sum_{i=c_0}^{c_T} \Phi_{pi} \tilde{\tilde{s}}_i\right\} \gtrless 0 \qquad (5.20)$$

The exponential argument is denoted by $\mathcal{A}$. Expanding (5.20)

$$\tilde{\tilde{\mathcal{A}}}_p = \left\{\left(\Phi_{pp} s_p - \Phi_{pp} \tilde{\tilde{s}}_p\right) + \sum_{i=c_0}^{p-1} \left(\Phi_{pi} s_p - \Phi_{pi} \tilde{\tilde{s}}_i\right) + \sum_{i=p+1}^{c_T} \left(\Phi_{pi} s_p - \Phi_{pi} \tilde{\tilde{s}}_i\right)\right\} \gtrless 0 \qquad (5.21)$$

This equation suggests that at any point p, the response corresponds to the summed up differences between its own response and the weighted functional distances of the surrounding stimuli. Expressing the summation limits in terms of the relative distances with respect to the point p

$$\tilde{\tilde{\mathcal{A}}}_p = \left\{ \left(\Phi_{pp} s_p - \Phi_{pp} \tilde{\tilde{s}}_p \right) + \sum_{\Delta i = p - c_0}^{1} \left(\Phi_{p-\Delta i} \left(s_p - \tilde{\tilde{s}}_{p-\Delta i} \right) \right) + \sum_{\Delta i = 1}^{c_T - p} \left(\Phi_{p+\Delta i} \left(s_p - \tilde{\tilde{s}}_{p+\Delta i} \right) \right) \right\} \gtrless 0 \tag{5.22}$$

Let signal s be extended such that the number of neighbouring elements on both sides of p are equal. The extension procedure will be discussed in detail in the next chapter. With this extension scheme, the summation limits can be replaced by their finite sizes. This leads to

$$\tilde{\tilde{\mathcal{A}}}_p = \left\{ \left(\Phi_{pp} s_p - \Phi_{pp} \tilde{\tilde{s}}_p \right) + \sum_{\Delta i = c_T}^{1} \left(\Phi_{p-\Delta i} \left(s_p - \tilde{\tilde{s}}_{p-\Delta i} \right) \right) + \sum_{\Delta i = 1}^{c_T} \left(\Phi_{p+\Delta i} \left(s_p - \tilde{\tilde{s}}_{p+\Delta i} \right) \right) \right\} \gtrless 0 \tag{5.23}$$

Without loss of generality the function Φ can be assumed even and symmetrical , i.e., ($\Phi_{(p-\Delta i)} = \Phi_{(p+\Delta i)}$). (5.23) then becomes

$$\tilde{\tilde{\mathcal{A}}}_p = \left\{ \left(\Phi_{pp} s_p - \Phi_{pp} \tilde{\tilde{s}}_p \right) + \sum_{\Delta i = 1}^{c_T} \left(- \tilde{\tilde{s}}_{p-\Delta i} + 2 s_p - \tilde{\tilde{s}}_{p+\Delta i} \right) \Phi_{p-\Delta i} \right\} \gtrless 0 \tag{5.24}$$

The term under summation can be readily identified as the convolution operation on the signal with a Laplacian mask -1 2 -1. From these results, we can understand that the enhancement mechanism is the result of using a filter function in the ratio form of (5.18) that leads to a Laplacian operation. (5.24) can therefore be written as

$$\tilde{\tilde{\mathcal{A}}}_p = \left\{ \left(\Phi(0) \left(s_p - \tilde{\tilde{s}}_p \right) \right) + \left(\nabla^2 s \left(\Delta i \right) \right) * \Phi(\Delta i) \right\} \gtrless 0 \tag{5.25}$$

Substituting (5.25) in (5.18), the true *ME* estimate becomes

$$\tilde{\tilde{s}}_p = q_p \exp\left(\Phi(0) \left(s_p - \tilde{\tilde{s}}_p \right) \right) \exp\left(\left(\nabla^2 s \left(\Delta i \right) \right) * \Phi(\Delta i) \right) \tag{5.26}$$

The properties of the Laplacian operation in relation (5.26) are studied in the next chapter. In what follows, the observations reported in the first part of this chapter are interpreted.

5.4. INTERPRETATIONS

It was shown in the first stage of analysis that the results of all the test classes that showed enhancement effects reduce to a common form (5.18). In this form, the classes are characterized by the function ϕ. In the absence of restoration, the equations (5.4) and (5.7.1) show that the parameter λ forms the characteristic function. In presence of restoration, the characteristic function is jointly determined by the point spread function h and λ as shown by (5.12.10), (5.17.1) and (5.7.1). These characteristic functions when present in the common form (5.18), lead to a Laplacian operation as shown by (5.26). It follows therefore that the enhancement effects shown by the test classes 1 or 2 or 3 are the direct consequences of the hidden Laplacian operation. The Laplacian operator in (5.26), as will be shown in the next chapter, in general, is not a sensitive edge enhancement function. This is because, the characteristic function has a significant influence on the Laplacian operation. However, under specific conditions, the operator shows increased sensitivity. For example, with a point spread function of 3×3 size, the Laplacian operator has a maximum sensitivity. These issues are discussed in the next chapter. The increased sensitivity of the Laplacian operator is the reason for the enhanced result in Fig.4.11.(d). The prominance of the features with increasing iterations as observed in Figs. 4.17(a),(b) and (c) follows directly from equation (5.26). As iterations proceed, the first exponential term tends to 1 because, the estimate approaches its true limit and the exponential argument tends to 0. As a result, the influence of the sensitive Laplacian operation is dominant in the estimate.

The above interpretations refer to the classes 2 and 3, where the influence of the restoration process is a matter of concern. The test class-1 is also susceptible to similar arguments because of the common form and the enhancement mechanism. But unlike the point spread function in (5.17.1), the type of influence of ϕ on the Laplacian operation is

not clearly known for this test class. Further, the parameter β cannot be overlooked in the interpretations, because the final estimate that showed the Mach band characteristics is the result of relaxation, obtained with

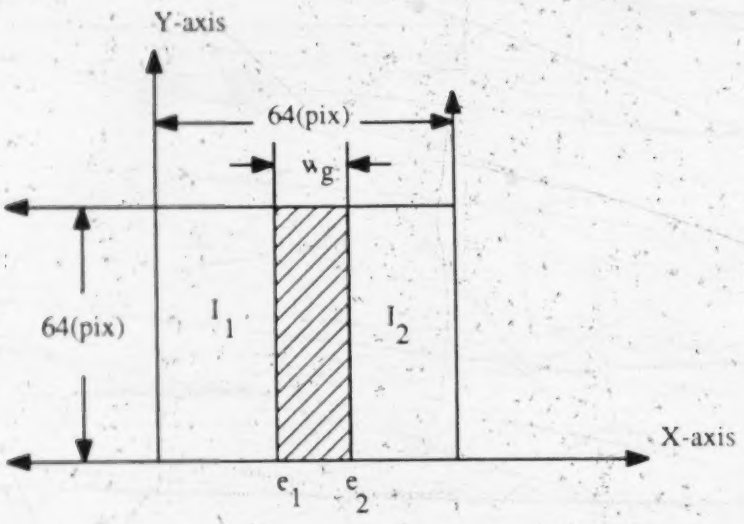
$$\tilde{f}_j^{(n)} = \sum_{i=0}^{n} \beta_p^{(i,n)} \tilde{f}_j^{(i)} \tag{5.27}$$

where

$$\beta_p^{(i,n)} = \left(1-\beta^{(n)}\right)\left(1-\beta^{(n-1)}\right)\ldots\ldots\left(1-\beta^{(i+1)}\right)\beta^{(i)} \tag{5.27.1}$$

with $\tilde{\tilde{f}}_j^{(i)}$ given by (5.26) and with f replacing the signal s. For the same reasons, the association of the entropy measure with the appearance of the bands in the estimate are not predictable. In short, the interpretations for this class need more investigations.

The common form (5.18) shows strong similarities to Mach's model. Mach's model was suggested as a possible neural network by Mach himself to account for the observed Mach band behaviour [80]. The model however remains obscured, perhaps because it was suggested before electro physiological evidences became available. The following chapter investigates Mach's model in detail, to explore its possible links with the *ME* estimation procedure.



gradient area w_g gradient width (pixels)

intensity slope = $(I_2 - I_1) / w_g$

Fig. 5.1 Test features of the pattern in Fig.4.18(a)

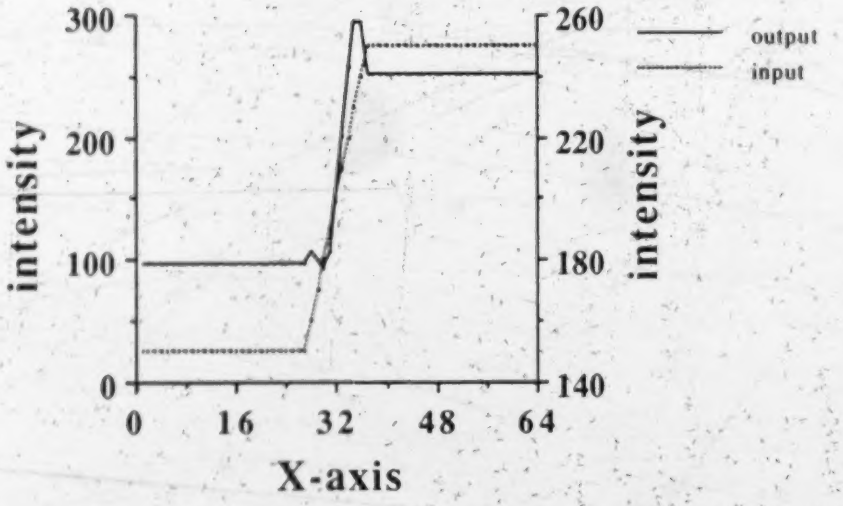
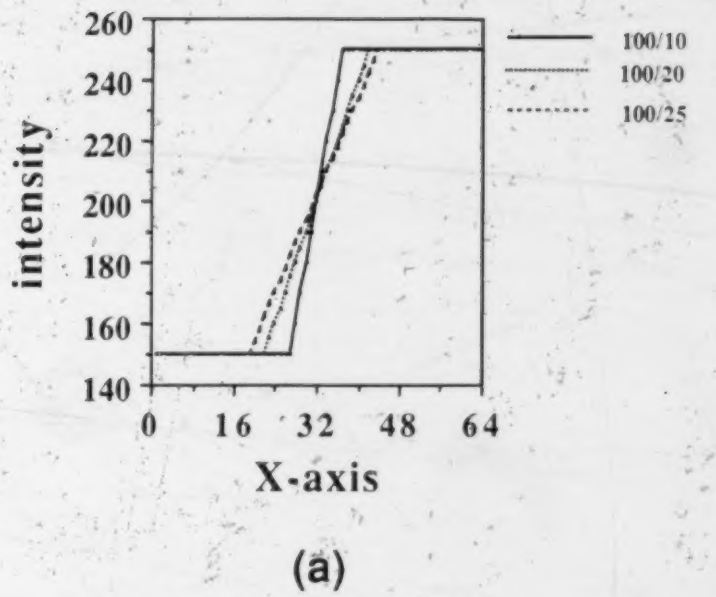
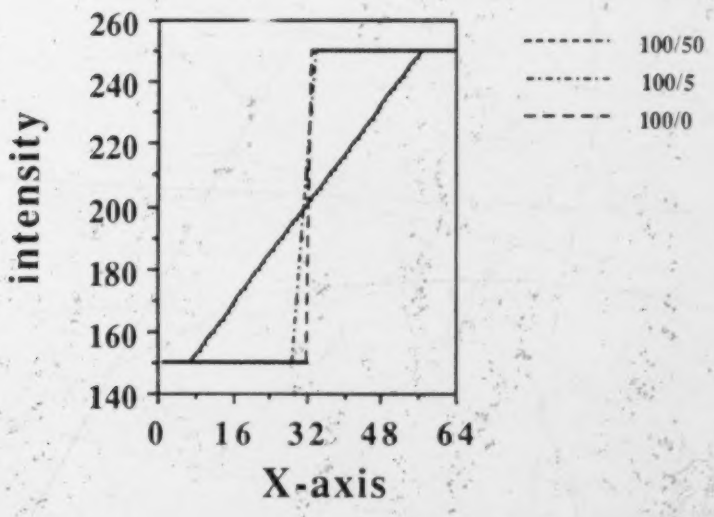


Fig. 5.2 The 1 - D intensity profile of the ME estimate of the test pattern

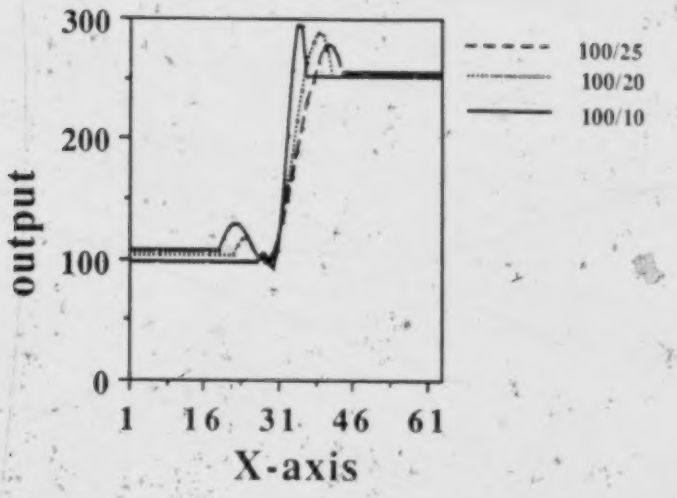


(a)

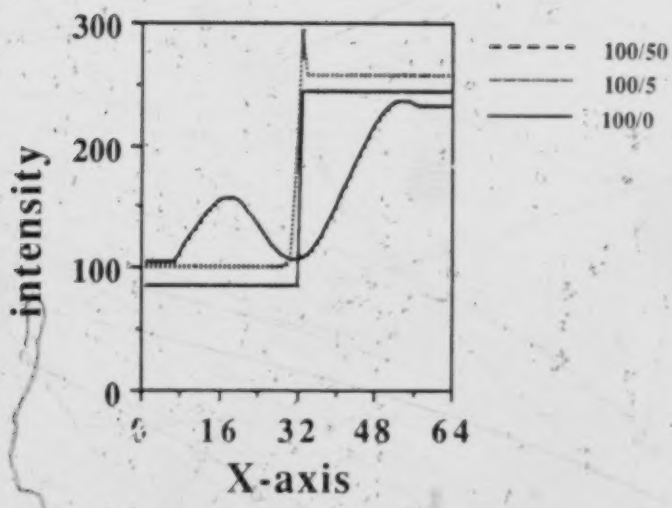


(b)

Fig. 5.3 A family of input test patterns with varying intensity gradient widths



(a)



(b)

Fig. 5.4 *ME* response to varying widths

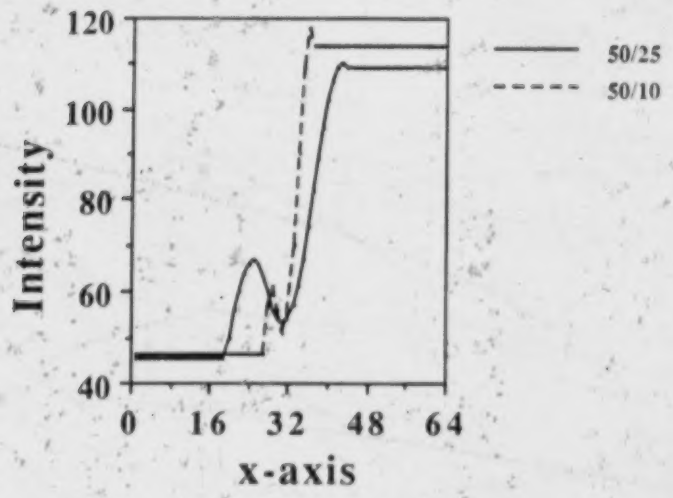


Fig. 5.5 *ME* response to illuminance changes

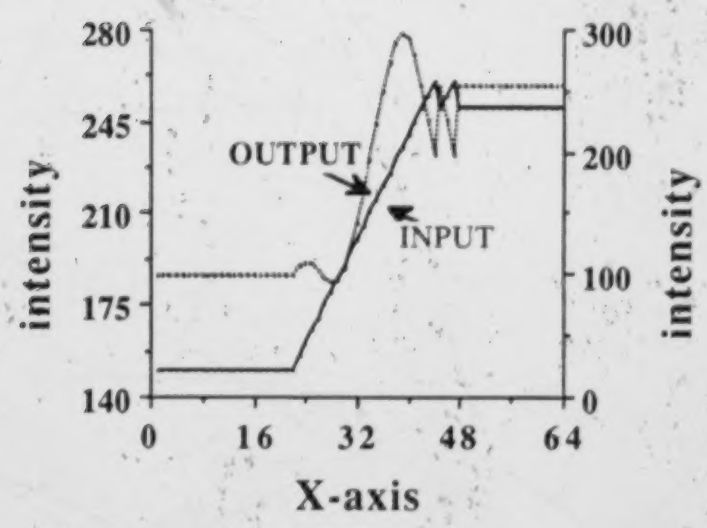


Fig. 5.6 A sensitivity test pattern and the ME response

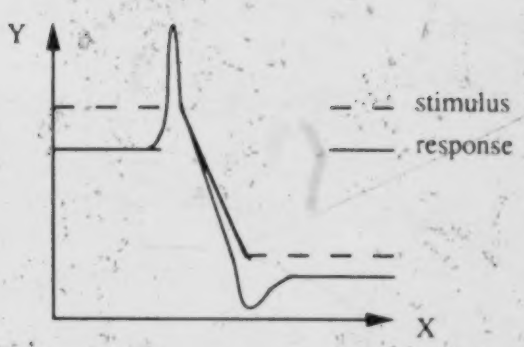


Fig. 5.7 Characteristics of Mach bands from psycho - physical studies

(schematic diagram)

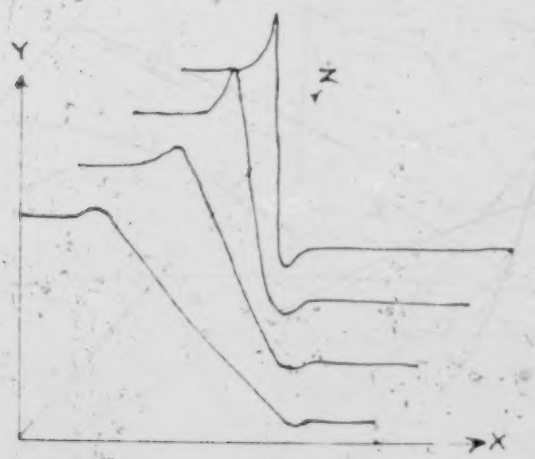


Fig. 5.8 Characteristics of Mach bands for varying widths from psycho - physical studies

(schematic diagram)

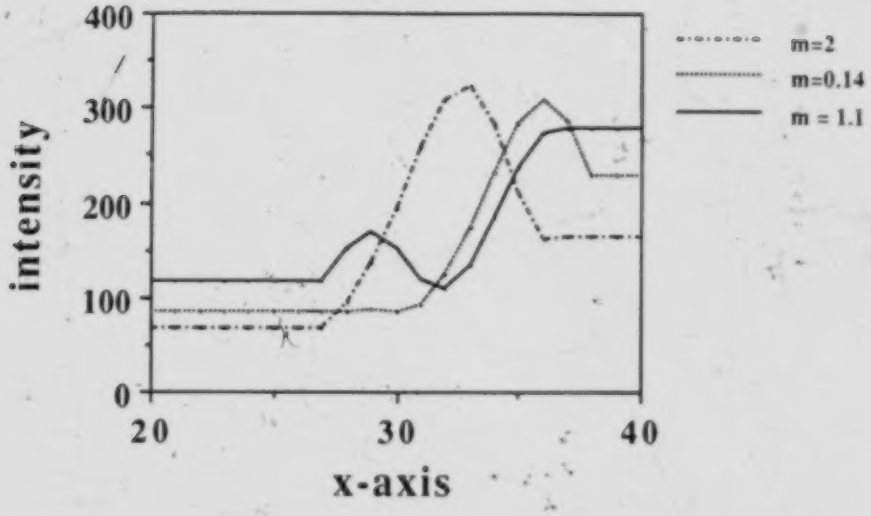


Fig. 5.9 Influence of β variations on the *ME* response to the tesp pattern in Fig. 5.1(a)

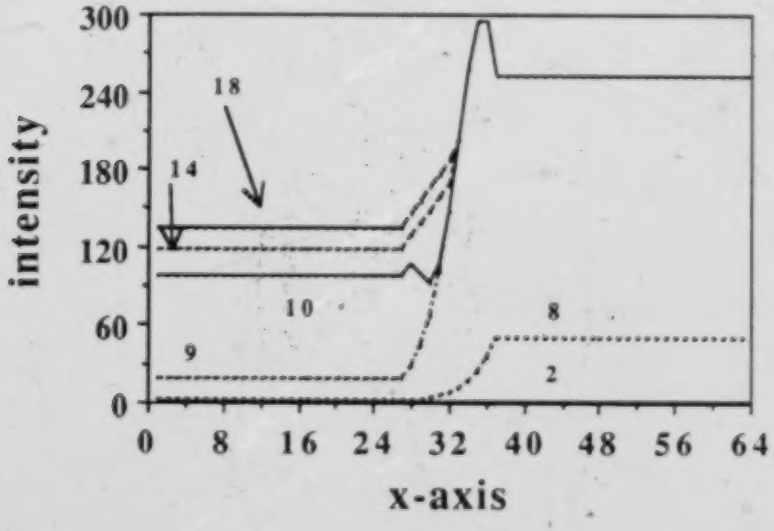


Fig. 5.10 The iterative sequence of the test pattern in the *ME* estimation

CHAPTER 6

STUDIES ON THE MACH'S BIOLOGICAL MODEL OF VISUAL ENHANCEMENT

The mathematical similarities of the *ME* estimation procedure with Mach's model need further investigations because, the results evolve from two different perspectives: computational and biological. Further, the obsoleteness of Mach's model, its appearance in connection with the *ME* estimation and the striking similarities with psycho-physical characteristics all need a satisfactory link for a better understanding. This chapter investigates the links by analysing Mach's model. The ideas arising out of this attempt are extended to the possibilities of using the *ME* method for edge detection and enhancement applications.

6.1. STUDIES ON MACH'S MODEL

Following the discovery of the Mach band phenomenon and the studies on psycho-physical characteristics, many neural network models were proposed, including Mach's model. These early models fall into the category of stimulus dependent (linear and non-linear) type [80]. Mach's model, in particular, is a non-linear, stimulus dependent model. When the electrophysiological studies on the eyes of Limulus (a marine arthropod) demonstrated the lateral inhibition phenomenon (inhibition of each receptor upon its neighbor receptors on the retinal surface) as the underlying cause of the Mach band phenomenon, new models began to appear under the class of response dependent (simultaneous-linear) type. With the change in the model class from stimulus-dependent to response-dependent, Mach's model went obscured. The analytical results of chapter 5 suggests the possibilities for a 'hybrid' model. Striking similarities with psycho-physical characteristics lead to the interpretation of the *ME* approach as a response-dependent

Mach's model. Following the aim of this thesis work, attention is focused on the possible developments of the *ME* algorithm with Mach's model as the basis, rather than the development of a *ME* visual model.

6.1.1. Mach's model - an overview

In its proposed form (as reported by Ratliff and modified here for consitency notations) Mach's model is given by the relation

$$r_p = f_p \left[\frac{f_p \sum_{j=1}^{n} \phi(\delta_{jp})\, \Delta a_j}{\sum_{j=1}^{n} f_j \phi(\delta_{jp})\, \Delta a_j} \right] \qquad (6.1)$$

where, r_p is the neural activity discharged from a particular element p (neuron) the stimulus intensity f_p. Mach's model, as can be seen from (6.1), has foreseen the fundamental inhibitory mechanism of biological systems before conclusive evidences became available through electrophysiological studies. Therefore, the obsoleteness of Mach's model is not a matter of concern for the analytical studies in the following sections.

The numerator term of 6.1 is the excitatory influence due to the stimulus intensity alone and the denominator is the inhibitory influence at a particular receptor p due to the intensities from the neighbouring receptors. Mach believed that the function ϕ must be decreasing with increasing distances of the neighbouring receptors from p (this means that the elements nearest to one another exert a greater influence upon each other). The distances are denoted by the symbol δ in (6.1). The individual inhibitory influences exerted on any particular element are additive. However, the model is non-linear, because the resulting neural activity from any element is the product of the intensity on that element and the ratio of excitatory and inhibitory influences.

A simple distribution which illustrates Mach's model is the step pattern, such as the one encountered earlier in Fig. 5.1(a). The pattern satisfies the following inequalities at the

edge points e_1 and e_2. At e_1

$$f_i > f_p \text{ for } i>p. \tag{6.2}$$

The numerator term of (6.1) being a low constant, is less than the denominator term, satisfying the condition

$$\sum_{j=1}^{n} f_p \phi(\delta_{jp}) \Delta a_j < \sum_{j=1}^{n} f_j \phi(\delta_{jp}) \Delta a_j \tag{6.3}$$

The result is $r_p < f_p$. Hence an undershoot is expected at e_1. At e_2 we notice

$$f_j < f_p \text{ for } j < p \tag{6.4}$$

leading to the relation

$$\sum_{j=1}^{n} f_p \phi(\delta_{jp}) \Delta a_j > \sum_{j=1}^{n} f_j \phi(\delta_{jp}) \Delta a_j \tag{6.5}$$

The condition (6.5) implies that $r_p > f_p$, which implies an overshoot at e_2. An example of this result is shown in Fig. 6.1 with the function $\phi(\delta_{pj})$ given by

$$\phi(\delta_{pj}) = \frac{1}{|\delta_{pj}|} \tag{6.6}$$

considering only the first 16 neighbouring elements. δ_{pj} denotes the distance between the receptor elements p and j. The important feature to be noted is that the response is non-linear with the amplitude of the overshoot higher than that of the undershoot. This asymmetry, as noted earlier, is one of the important features of the Mach bands observed in the experimental studies.

6.1.2. The underlying mechanism of enhancement

Since Mach's model falls directly under the general form (5.18), the derivation

steps of this section follow closely the ones reported under 5.3.4, but in a different context. The analysis is carried out in 1-D for convenience. Considering the ratio term R_p of (6.1) for a particular element p we can write

$$R_p = \frac{B_p}{I_p} = \frac{\sum_{j=1}^{n} I_p \, \phi(\delta_{jp}) \, \Delta a_j}{\sum_{j=1}^{n} I_j \phi(\delta_{jp}) \, \Delta a_j} \qquad (6.7)$$

The term f_p in (6.1) has been replaced by I_p in (6.7) for generality. Let us assume that all receptors have an equal surface area Δa. Expressing (6.7) as an inequality relation and denoting it with E_p, we obtain

$$E_p = \left[I_p \sum_{j=1}^{n} \phi(\delta_{pj}) \, \Delta a - \sum_{j=1}^{n} I_j \, \phi(\delta_{pj}) \, \Delta a \right] \lesseqgtr 0 \qquad (6.8)$$

Let us express (6.8) in terms of the responses from the receptor area of the particular element and its neighbouring area:

$$I_p \left[\left(\sum_{j=1}^{p-1} \phi(\delta_{jp}) + \sum_{j=p+1}^{n} \phi(\delta_{jp}) + \phi(\delta_{pp}) \right) \Delta a - \left(\sum_{j=1}^{p-1} I_j \phi(\delta_{jp}) + \sum_{j=p+1}^{n} I_j \phi(\delta_{jp}) + I_p \phi(\delta_{jp}) \right) \Delta a \right] \gtrless 0 \qquad (6.9)$$

Grouping the receptors yields

$$\left[\sum_{j=1}^{p} \left(I_p \phi(\delta_{pj}) - I_j \phi(\delta_{pj}) \right) \Delta a + \sum_{j=p+1}^{n} \left(I_p \phi(\delta_{pj}) - I_j \phi(\delta_{pj}) \right) \Delta a \right] \gtrless 0 . \qquad (6.10)$$

The self inhibitory terms get cancelled in (6.10). Replacing j in terms of its relative distance δ_{pj} from p results in

$$\left[\left\{ \sum_{\delta_{jp}=p-1}^{1} \left(I_p - I_{p-\delta_{jp}} \right) \phi(\delta_{jp}) \, \Delta a \right\} + \left\{ \sum_{\delta_{pj}=1}^{n-p} \left(I_p - I_{p+\delta_{pj}} \right) \phi(\delta_{jp}) \, \Delta a \right\} \right] \gtrless 0 \qquad (6.11)$$

where

$$\delta_{jp} = x_p - x_j \qquad (6.11.1)$$

$$\delta_{pj} = x_j - x_p \qquad (6.11.2)$$

The function $\phi(\delta_{pj})$ is assumed to be real and even so that $\phi(\delta_{pj}) = \phi(\delta_{jp})$. For analytical purposes, let us extend (virtual extension) the neighbouring receptor fields to n without affecting any of the receptors under the active region. The extension procedure is discussed in detail in the next section. The basic idea behind the extension for (6.11) is to make the summation limits equal with respect to p. Equation (6.11) then becomes,

$$\left[\sum_{\delta_{jp}=1}^{n} \left\{ - I_{p-\delta_{jp}} + 2I_p - I_{p+\delta_{jp}} \right\} \phi\left(\delta_{jp}\right) \Delta a \right] \gtrless 0 . \qquad (6.12)$$

(6.12) has a Laplacian term often denoted as

$$\left[- I_{p-\delta_{jp}} + 2I_p - I_{p+\delta_{jp}} \right] \equiv \nabla^2 I_p \left(\delta_{jp}\right) . \qquad (6.13)$$

(6.13) is a convolution operation of I_p and the Laplacian mask [-1 +2 -1]. Substituting (6.13) back in (6.8) yields

$$E_p = - \sum_{\delta_{jp}=1}^{n} \nabla^2 I_p \left(\delta_{jp}\right) \phi\left(\delta_{jp}\right) \gtrless 0 . \qquad (6.14)$$

Subtracting 1 from both sides of (6.7) and substituting for Ep from (6.14), we can arrive at an alternative expression form for Mach's model:

$$B_p = \frac{I_p \sum_{\delta_{jp}=1}^{n} \nabla^2 I_p\left(\delta_{jp}\right) \phi\left(\delta_{jp}\right)}{\sum_{j=1}^{n} I_j \phi\left(\delta_{jp}\right)} + I_p . \qquad (6.15)$$

(6.15) shows that the underlying mechanism of the model is Laplacian with the same form derived in (5.26) for the common form of the *ME* estimation procedure.

6.1.3. Boundary correction procedure

The derivation steps above made use of an extension procedure to equalize the number of neighbours for an element p. Extension becomes necessary when

$$p- \delta_{jp} < 1 \text{ and } p+ \delta_{jp} > n. \tag{6.16}$$

Under this situation, the functional invariance of the Laplacian operation is satisfied in (6.13) if

$$I_{p+\delta_{jp}} \overset{>}{\leq} I_p \text{ if } p-\delta_{jp} < 1 \tag{6.17.1}$$

$$I_{p-\delta_{jp}} \overset{>}{\leq} I_p \text{ if } p+\delta_{jp} < n. \tag{6.17.2}$$

(6.17.1) and (6.17.2) are realized in (6.13) if

$$I_{p-\delta_{jp}} = I_{p+\delta_{jp}} \text{ if } p-\delta_{jp} < 1. \tag{6.18.1}$$

$$I_{p+\delta_{jp}} = I_{p-\delta_{jp}} \text{ if } p+\delta_{jp} > n \tag{6.18.2}$$

This is indeed mirror extension, one of the commonly encountered boundary correction techniques in image convolution operations. Instead of extending the sequence, correction terms can be added to (6.7) depending on the number of neighbours in the function ϕ. That is

$$\left(I_{p+\delta_{jp}}\right) \phi\left(\delta_{jp}\right) \text{ if } p-\delta_{jp} < 1 \tag{6.19.1}$$

$$\left(I_{p-\delta_{jp}}\right) \phi\left(\delta_{jp}\right) \text{ if } p+\delta_{jp} > n \tag{6.19.2}$$

For a given p and δ_{pj}, the correction terms (6.19.1) and (6.19.2) can be added to (6.7).

6.2. LAPLACIAN OPERATOR IN THE MODEL

Rewriting the enhancement mechanism of the model given by (6.14)

$$R_p = \sum_{\delta_{jp}=1}^{n} \nabla^2 I_p\left(\delta_{jp}\right) \phi\left(\delta_{jp}\right) \overset{>}{\leq} 0 . \tag{6.20}$$

where **n** is the order of ϕ (1/2 (number of coefficients-1). (6.20) can be expressed in a simple form

$$R_p = L_{\phi n} I_{pn}^T \gtrless 0 \qquad (6.21)$$

where

$$L_{\phi n} = [-\phi_n \; -\phi_{n-1} \cdots \phi_{-1} \; \phi_0 \; \phi_1 \cdots \phi_{n-1} \; \phi_n] \qquad (6.21.1)$$

with

$$\phi_0 = 2\sum_{i=1}^{n} \phi_i. \qquad (6.21.2)$$

Also

$$I_{pn} = [I_{p-n} \; I_{p-n+1} \cdots I_{p-1} \; I_p \; I_{p+1} \cdots I_{p+n-1} \; I_{p+n}] \qquad (6.21.3)$$

The coefficients of ϕ are in the form $L_{\phi n}$. As can be seen from (6.21.1) and (6.21.2), irrespective of the type and order of ϕ, the elements of L always satisfy the condition

$$\sum_{i=-n}^{+n} \phi_i = 0. \qquad (6.22)$$

Therefore $L_{\phi n}$ can be expressed as a Laplacian in (6.21) resulting in

$$R_p = \nabla_{\phi n}^2 I_{pn}^T \gtrless 0 \qquad (6.23)$$

When the operation over the whole image is considered, we obtain

$$R = \nabla_{\phi n}^2 * I. \qquad (6.24)$$

Alternatively

$$R = \nabla^2 (I * L_{\phi n}) \qquad (6.25)$$

The symbol * denotes convolution operation. Because of convolution, the Laplacian operation in the model is strongly influenced by the type and the order of ϕ. With the specific one neighbour case **n** = 1, (6.24) becomes

$$R = \nabla_{\phi n}^2 * I \equiv \nabla^2 (I * \phi) \qquad (6.26)$$

which is a standard Laplacian operation. With increasing order of the filters, there is a problem of decreased edge sensitivity due to the convolution operation. But using (6.20), we can infer that the fundamental component (the order 1 component) remains hidden in (6.20). By subtracting the higher order Laplacian responses, it is possible to remove the fundamental. Such an approach is however not practical because of the significance of the convolution operation in noise reduction. In presence of noise (zero mean and additive), a higher order increases the number of summation terms and hence noise is reduced by averaging. Let the input be

$$g_p = I_p + \varepsilon_p. \tag{6.27}$$

Replacing I_p by g_p in (6.21)

$$R_p = L_{\phi n}(I_{pn} + \varepsilon_{pn})^T \gtrless 0 \tag{6.28}$$

where $L_{\phi n}$ and I_{pn} both take the same form as in (6.21.1)-(6.21.3). The noise terms ε_{pn} are given by

$$\varepsilon_{pn} = [\varepsilon_{p-n}\ \varepsilon_{p-n+1} \cdots\cdots \varepsilon_{p-1}\ \varepsilon_p\ \varepsilon_{p+1} \cdots\cdots \varepsilon_{p+n-1}\ \varepsilon_{p+n}] \tag{6.29}$$

(6.28) and (6.29) shows that with the higher order filter the noise terms (0 mean) effectively get cancelled thereby reducing its effect on R_p. The standard deviation σ of the noise in the result varies inversely with the order of the filter ϕ, following an error reduction operation by averaging .

6.3. IMPLICATIONS

Edge enhancement is an important operation for many image processing applications. Laplacian operation is one of the popular edge enhancement operations (with no reference to edge direction), best understood as the convolution of an image with a Laplacian mask [28], [82], [83]. The masks are often generated with coefficients satisfying

the fundamental condition in (6.22). Examples of such masks are

$$\begin{matrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{matrix} \qquad \begin{matrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{matrix} \qquad \begin{matrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{matrix} \tag{6.30}$$

(a) (b) (c)

The Laplacian operation in Mach's model on the other hand has no external representation as in (6.30). It is inherent in the model itself in the form (6.15). Besides Laplacian, there are many other enhancement operations scattered throughout in the image processing literature.

6.3.1. Unsharp - masking Vs Mach's model

With a normalized Φ, Mach's model can be written in the form

$$\frac{B_p}{f_p} - 1 = \left[\frac{f_p - \sum_{j=1}^{n} f_j \Phi_{jp}}{\sum_{j=1}^{n} f_j \Phi_{jp}} \right] \tag{6.31}$$

After simplification we obtain

$$f_p - f_{pl} = f_{pl}\left(\frac{B_p}{f_p} - 1\right) \tag{6.32}$$

where

$$f_{pl} = \sum_{j=1}^{n} f_j \Phi_{jp} \tag{6.32.1}$$

Multiplying (6.32.1) by a constant c and adding f_{pl} on both sides we get the relation

$$f_{pl} + c(f_p - f_{pl}) = c f_{pl}\left(\frac{B_p}{f_p} - 1 + \frac{1}{c}\right) \tag{6.33}$$

The left hand side term can be recognized as an unsharp masking operation , a well known technique of edge enhancement used in electronically scanned images [84]. It is a simple linear technique, where an image is subtracted from its blurred version in order to

accentuate the edges. f_{pl} is a low pass filtered image intensity at pixel p and c is a boost factor in this technique. Replacing the left hand side term by f_{pum} with the suffx 'um' denoting the unsharp masking operation.

$$\frac{f_p}{c}\left(\frac{f_{pum}}{f_{pl}} - 1 + c\right) = B_p \qquad (6.34)$$

With c = 1, it can be seen that, the unsharp masked image and Mach's mdoel differ by a mutiplication factor. The multiplication factor is the ratio formed by an image and its filtered (low-pass in this case) version. Using (6.15), (6.34) can be expressed in terms of a Laplacian operation (with c = 1):

$$f_{p(um)} = f_p\left(\sum_{\delta_{jp}=1}^{n} \nabla^2 I_p(\delta_{jp})\,\alpha(\delta_{jp}) + f_{pl}\right) \qquad (6.35)$$

(6.35) is the unsharp masking operation using the Laplacian operator of Mach's model.

6.3 2. Mach's model and *ME* edge enhancement

The analysis of Mach's model has shown that, a filter or a point spread function when used in the form of Mach's model, leads to Laplacian operation. The Laplacian mask is of type (6.30)-(a) for 2-D images and the operator is given by (6.21.1) and (6.21.2). The performance of Mach's model in edge enhancement is illustrated with an example in Fig. 6.2. The image B_j; j = 1, ..., n transformed by Mach's model, as shown in Fig. 6.2(b) has been subject to a thresholding operation given by

$$E_j = 1 \quad \text{if } B_j \geq \theta \qquad (6.35.1)$$
$$E_j = 0 \quad \text{if } B_j < \theta \qquad (6.35.2)$$

where θ is the thresholding parameter, fixed by the mean of the image B. The thresholded image is shown in Fig.6.2(c). With a slightly lowered threshold, the edge image looked like Fig. 6.2 (d).

The image results are only an illustration to show the performance of Mach's model with real images. Significant information can be obtained by evaluating its performance with the test patterns introduced in chapter 5. The response of Mach's model to the stimulus patterns with varying widths is shown in Figs. 6.3(a)and (b). The Laplacian operator that led to this image has followed the form (6.21.1) with ϕ as the 3 x 3 Gaussian filter mask. The results of the *ME* method, explained in chapter 5 are shown in Fig. 6.3(c) and (d). The *ME* response is the result of a relaxation procedure under specific parameter conditions. Comparison of both responses with the psycho-physical characteristics shown in Fig. 5.8 shows that the *ME* response follows more closely the expected results. The significance of the response dependent models is understood from this test result. The importance of suitable parameter conditions in obtaining the expected results was already discussed in chapter 5. All these evidences suggest that the *ME* method in the parameter relaxation approach is a possible response dependent biological visual model. It is however not conclusive without exploring the type and influence of the parameter conditions of the model on its response. The results also suggest that the *ME* method can be used for edge detection and enhancement applications in computational vision studies. Recognising the fact that Fig. 4.17(c) is the product of an image restoration process, it can be considered as a true image distribution, although the Laplacian effects are dominant in the result. Using Mach's model, the edge features of the restored result can be enhanced as shown in Fig. 6.3. The enhancement of the iris is the notable feature in the result. More studies are however necessary to explore the potential of using the *ME* method for edge detection and enhancement and applications. With its foreseen potentials, attention is focused on the implementation issues in the next chapter.

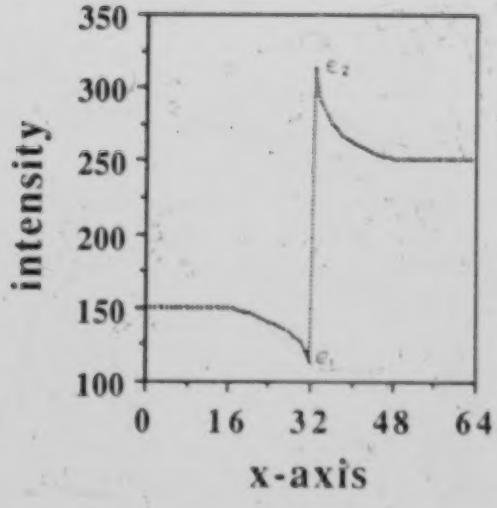


Fig.6.1 Response of Mach's model to the test pattern in Fig. 5.1(a)

(a) (b)

(c) (d)

Fig.6.2 Edge enhancement performance of the Mach's model

(a) Input image; (b) Edge enhanced image

(c) and (d) Mean thresholded images

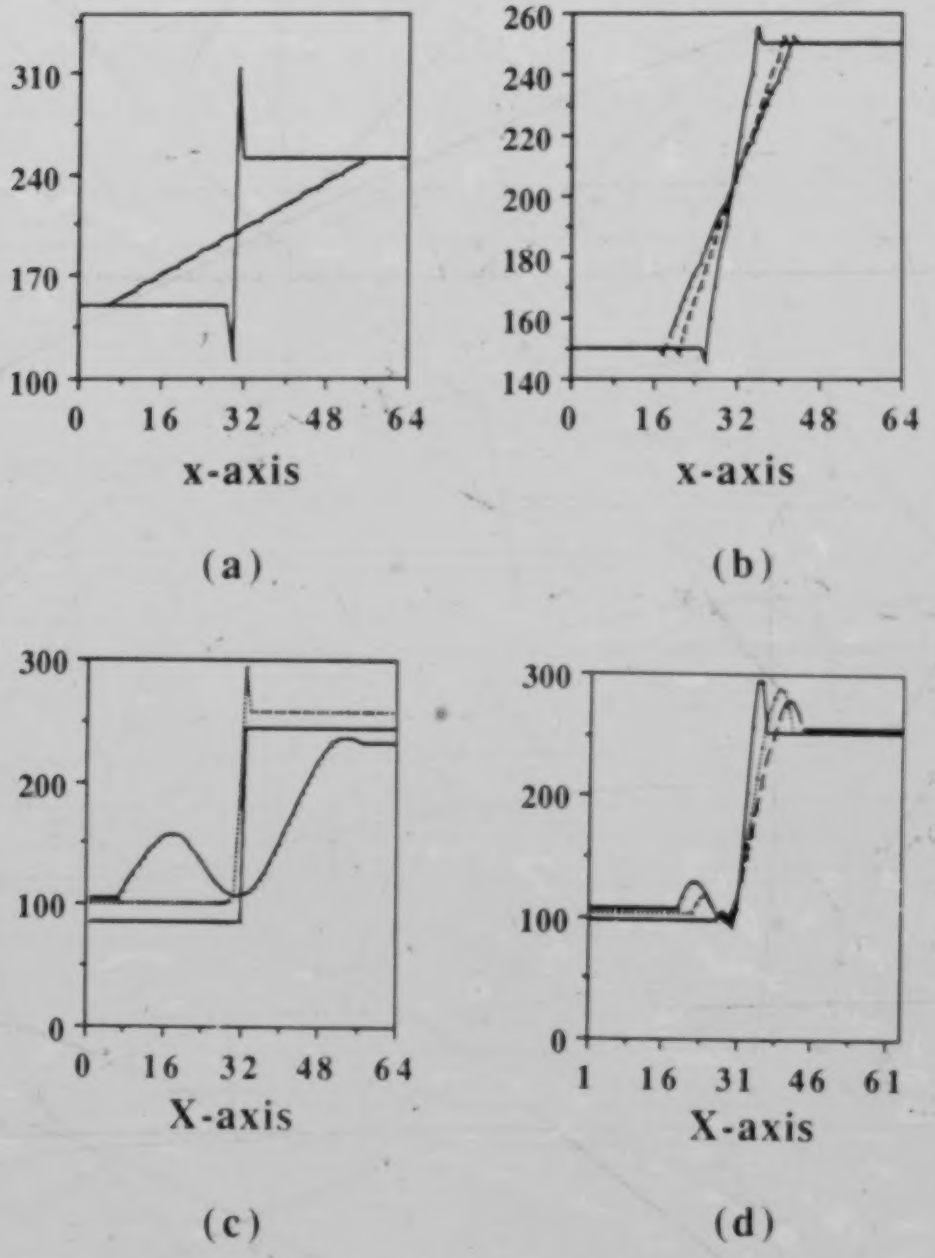
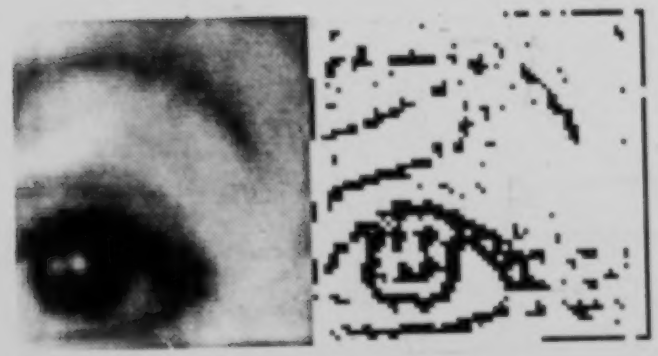


Fig.6.3 Response of Mach's model Vs *ME* estimation procedure to stimulus patterns with varying widths

Response of (a) Mach's model; (b) *ME* estimation procedure



(a) (b)

Fig.6.4 Edge performance of the *ME* method - an example

(a) Input image; (b) Edge enhanced image (thresholded)

CHAPTER 7

PARALLEL IMAGE CONVOLUTION *ME* IMAGE DECONVOLUTION - AN IMPLEMENTATION STUDY

Computationally, the problem of obtaining a convergent solution for the *ME* algorithm is the same as that of any constrained optimization algorithm. A possible solution was proposed in chapter 3 in the context of image processing applications. In this scheme, the solution is obtained by controlling two parameters of a solution equation under a relaxation scheme, with the convergence limits set by the entropy maxima. Despite this simplification, it was found in chapter 4 that the algorithmic time complexity of its application class was as high as $O(N^4)$, where N^2 is the image size, typically ranging from 256 x 256 to 1024 x 1024 (square image is assumed for discussion convenience). For the meaning of the O- notation of time complexity, readers are referred to the 'list of symbols' section of this thesis. Sequential algorithms are impractical for large image sizes. In fact, to reduce the time complexity of the sequential algorithm for test studies, the problem size had to be restricted to 64 x 64. Parallel processing implementation offers the potential to reduce time complexity, by having multiple processors cooperating to solve a given computational problem without restricting the problem sizes. The gain in performance relies heavily on the effective parallelization the algorithms to map them efficiently onto parallel hardware. This chapter explores the problems and benefits in parallelizing the *ME* algorithm using the parameter relaxation approach. Studies are carried out for a specific application class of image convolution and deconvolution using the Myrias SPS-2 parallel computer.

7.1. THE *ME* ALGORITHMS - A REVIEW

The computational problem of the *ME* method is to find a set **x** of independent

variables x_i, i = 1,.....,N that would yield the absolute maximum of a cost or objective function $J(x_i)$. In image processing applications, the independent variables are the individual pixels. The objective function for the present study is the standard maximum entropy method expression in (2.7) or (2.16) under a Bayesian approach. In the following section, the general means of obtaining a solution for the maximization problem are discussed including the parameter-controlled relaxation procedure proposed in this thesis. All discussions refer to image processing applications.

7.1.1. Dynamic programming techniques

In general, the maximization of the objective function in (2.7) involves finding a solution $\hat{f}^*$ by satisfying a set of $m + 1$ constraints using $m+1$ Lagrange multipliers. The number of constraints depends upon the type of application and the solution approach. In its simpler form the objective function is given by

$$J(\mathbf{p}^*) = S(\mathbf{p}^*) + \mu\left(\sum_{i=1}^{n} \hat{p}_i^* - 1\right) + \sum_{k=1}^{m} \lambda_k\left(\sum_{i=1}^{n} A_{ki}\hat{p}_i^* - d_k\right) \qquad (7.1)$$

where $n = N^2$ in the present context and m is the measured image size. Earlier attempts approached this problem by solving (7.1) as a set of m non-linear equations [11]. The problem with this approach is that the number of equations increases with increasing problem sizes. Subsequent attempts therefore made use of the single constraint *ME* method explained under section 2.3.4. The computational problem for this scheme simplifies to the unconstrained maximization of

$$J(\hat{f}) = S(\hat{f}) - \lambda Q(d\,\hat{f}) \qquad (7.2)$$

with a single Lagrange mutiplier and a single constraint in (2.15). Most of the solution techniques for (7.2) explained in chapter 2, fall into the class of dynamic programming techniques with iterative improvements. The optimality conditions of the solution with

respect to the local or global maximum conditions are hard to define in these search procedures. This often results in convergence problems.

7.1.2. Simulated annealing by stochastic relaxation

Simulated annealing techniques are designed to overcome the optimality problems with local and global convergence [72], [73]. In these techniques, the problem is stated as a minimization of an 'energy' function $E(\mathbf{s})$, defined on a finite set of states $\mathbf{S}$. The solution technique makes use of a relaxation scheme, where an initial choice of a state is refined and updated in the successive stages of iterations. The annealing algorithm generates a sequence of states based on a stochastic decision process. The basic idea of annealing is to allow the possibility of getting out of the trap of the local optimum. The decision scheme is based on the probability distribution

$$p_i = \exp\left\{\frac{-\mathrm{Max}[\Delta E, 0]}{T_i}\right\} \qquad (7.3)$$

with $\Delta E = E(\mathbf{s}_i') - E(\mathbf{s}_i)$ and T_i is the 'temperature' control parameter. In the beginning of the simulated annealing process, T is very high and the probability of accepting the new state $\mathbf{s}_i'$ is very large. This means that if the state is a local minimum, there is a possibility that it can get out of that state. During the iteration process, the temperature parameter is gradually lowered using a prescribed schedule for a guranteed convergence. When the temperature has become sufficiently lowered, the state will arrive at the global minimum region. In practice, the annealing process is found to be very slow. Geman and Geman have recently proposed simulated annealing algorithm for image restoration [25]. It follows a Bayesian approach to find a maximum *a posteriori* estimate of the original image modelled using Gibbs distribution.

7.1.3. Parameter-controlled relaxation

The procedure proposed in this thesis work falls under this category. Along the same line of reasoning followed in simulated annealing, the problem is approached as a physical system $\mathbf{S}$ to be optimized by a 'treatment' process in order to reach the ground state. The treatment process consists of a controlled relaxation of the system between its two limiting states. The controlling is done by the algorithm using a parameter β under the relaxation scheme in (3.35). The algorithm controls the sequences of state changes by controlling the behaviour of the system in response to external conditions. (for analytical details readers are referred to section 3.4 of chapter 3). The mechanism of the algorithm is briefly outlied below:

Initially, the system is at its stable state given by

$$\mathbf{s}_0 \leftarrow \tilde{\mathbf{f}}^{(0)} = \left\{ f_j^{(0)} \right\} ; \quad j = 1, \cdots\cdots n. \tag{7.4}$$

The initial estimates are given by (3.11). The behaviour of the system is characterized by the relation

$$\mathbb{B}(\mathbf{s}_0) < S^{(0)} q_a^{(0)}. \tag{7.4.1}$$

$S^{(0)}$ is the entropy of the state $\mathbf{s}_0$ given by (3.2) and q_a is the average scale parameter given by the logarithm of (3.2.3). The system is forced to leave its initial state by the external constraint(s), characterized by the parameter λ and σ^2. Depending on the strength of these external constraint parameters, the states get changed to

$$\tilde{\mathbf{s}}_i \leftarrow \tilde{\mathbf{f}}^{(i)}. \tag{7.5}$$

The behaviour of the states is typically characterized by

$$\mathbb{B}(\tilde{\mathbf{s}}_i) > S^{(i)} q_a^{(i)}. \tag{7.5.1}$$

As the strength of the constraint increases, the state of the system tends to become unbounded with its behaviour characterized by the relation

$$\mathcal{B}(\tilde{s}_i) >> S^{(i)}\hat{q}_a^{(i)} \qquad (7.5.2)$$

To control the unbounded behaviour of the system , the algorithm attempts to 'relax' the states using the parameter $\beta^{(i)}$ given by (3.62). When the parameter λ varies depending on the constraint condition, β controls the states of the system accordingly. The relaxation process modifies the states to

$$\bar{s}_i \leftarrow \bar{\tilde{f}}^{(i)} \qquad (7.6)$$

changing the behaviour to

$$\mathcal{B}(\bar{s}_i) \gtrless S^{(i)}\hat{q}_a^{(i)} \qquad (7.6.1)$$

$S^{(i)}$ is given by (3.70.3). With the continued relaxation process, at some stage m, when the entropy becomes maximum, the system begins to obey the condition

$$\mathcal{B}(\bar{s}_m) \leq S^{(m)}\hat{q}_a^{(m)}. \qquad (7.7)$$

With the behaviour of the system characterized by (7.7), the increased strength of the external constraints has minimal influence on the system. The state of the system can be denoted as

$$\bar{s}_m \leftarrow \bar{\tilde{f}}^{(m)}. \qquad (7.7.1)$$

The condition implied by (7.7.1) is measurable by a standard convergence measure given by (3.79). The scaling of (7.7.1) is arbitrary, since the relaxation process has taken place without the input scaling information. With an external upper bound constraint, the process can be controlled for the given scaling requirements so that

$$\bar{s}_{sm} \leftarrow \bar{\tilde{f}}^{(sm)}. \qquad (7.7.2)$$

The iterative refinement of each state is controlled by the intermediate constraint in (3.4.1). The rate of convergence (number of iterations) can be controlled by adjusting the component K_λ in λ. There are no strict restrictions observed for this control.

7.2. PARALLEL IMPLEMENTATION

As indicated earlier, the development of a suitable concurrent algorithm constitutes a key step for parallel implementation. This algorithm is then mapped onto a parallel architecture for the best possible exploitation of the exhibited concurrency. The term *concurrent* is used in this context to express *potential parallelism* ; the actual parallelism is determined during the mapping step. Typically, for efficiency reasons, several concurrent activities may be grouped onto a single processing element, depending on the level of parallelism (granularity) of both the design and the target architecture. Envisageable target architectures may range from special purpose processors to general purpose parallel machines. Systolic arrays are popular representatives of the first category [85]: they are the most suitable for mapping highly regular problems (e.g. matrix operations). Most 'real life' problems however contain a significant irregular component which tends to clutter the border of such arrays, at the expense of the otherwise excellent space/performance ratio. In such cases approaches as proposed by [86], deriving the target VLSI architecture from the concurrent algorithm description, may prove to be the most efficient. There exists many different viewpoints in classifying the general purpose parallel machines [87], [88], [89]. In the traditional classification [87], the members in the second category, the general purpose machines, may range from SIMD (single instruction, multiple data) to MIMD (multiple instruction, multiple data). For example, the vector processor CYBER - 205 is a pipeline machine (a special class of SISD). The MISD (multiple instruction, single data) class, sometimes seen as dataflow machines, does not have any commercially accessible implementations yet, but is the object of high research interest. Assuming that Flynn's taxonomy is complete, the MIMD class of computers cover a wide range of modern architectures. It is therefore useful to subdivide the MIMD class further into 3 subclasses in terms of granularity: coarse-grained, medium-grained and fine-grained. The first subclass

consists of multi computers or distributed systems: fully fledged computers communicate via relatively low bandwidth channels such as local- or wide area networks. The second subclass consists of general purpose processors (typically 32 to 1024 CPUs), embedded in hierarchical memory schemes, and communicating via high performance data paths (bus). Hypercube type machines are popular representatives of this subclass. The subclass of fine-grained MIMD, also called massively parallel machines consists of vast amounts of very primitive processors (typically 1K to 64K units) communicating through very highly developed communication schemes. The Connection machine of Hillis is the best known representative of this subclass [90].

The following elaborations will rely on the Myrias SPS-2, a medium-grained MIMD machine consisting of 64 M68020 processors. It should be noted that this architecture is not necessarily the best imaginable for the types of parallelism that will be considered in this chapter.

7.2.1. Myrias SPS-2 parallel computing system

Myrias SPS-2 is a multiprocessor parallel computing system (with 64 processors in test configuration) with a *pardo* extension to Fortran and C [91]. The syntax of *pardo* is identical to that of a conventional DO loop in Fortran. Each iteration of a *pardo* loop is a separate task with its own memory space. When a 'parent task' executes a *pardo* instruction, a collection of 'child' tasks is created and the parent task is suspended until all of the child tasks complete. Child tasks inherit identical copies of the parent task's memory with only the iterator differing. Each task then executes independently. Since each child task executes within its own memory space, sibling tasks cannot affect each other's memory. Upon completion of the child tasks, their memory images are merged together to form a new memory image for the parent, wherafter, the parent task resumes execution. The assignment of tasks to processors is hidden from the user and is done dynamically by the operating system (The Parallel Application Management System (PAMS)). The

operating system manages all of the copying, storage, and merging of memory images. Appendix-F shows the sequence of activities in the Myrias *pardo* scheme.

Proper understanding of the merging rules is fundamental to the efficient and reliable use of the SPS system. They are briefly outlined below following the Myrias manual specifications [91]. The merging rules are such that if exactly one child task changes a location, then the new value appears in the parent's memory image. If no child stores a value in a location, that location in the parent memory state retains the value it had before the execution of the *pardo* statement. If more than one child stores a value in a location, and the last value stored is the same for each child, that location in the parent memory state receives that value. If more than one child stores a value in a location, and the last value stored is not the same for each child, that location in the parent memory state is undefined when the parent resumes execution. The wait time of the parent task during merging, the communication requirements of child tasks execution, task assignment to processors, merging operation all constitute system overhead. The following section explores the consequences of these issues by parallelizing an image convolution and a deconvolution algorithm.

7.2.2. The concepts of instruction and domain partitioning parallelism

Let us consider a program P operating on an input image domain $\mathbb{D}$. A program in general may be considered as a set of instructions $\{I_j\}$ with $I_j \in P$; $j = 1,\ldots, J$, operating on a data set $\mathbb{D}$ ($\mathbb{D} \subseteq \mathbb{D}$) denoted as

$$P_{\mathbb{D}} \equiv \{I_j\}_{\mathbb{D}} \text{ for } j = 1,\ldots,J. \qquad (7.8)$$

The program, when executed on a given computer is defined in terms of an execution set $\{I_s\}$ with an execution time T. The execution program can be expressed as

$$P_s^T|P \equiv \left| (I_{s_k})_{D_k \subseteq D}^{\tau_{s_k} \in T} \right| \text{ for } k = 1,....,K. \qquad (7.9)$$

The time interval for the execution of instruction k is denoted as $\tau_{sk} = (t_{osk}, \Delta t_{sk})$. t_{osk} is the time of execution initiation instruction k and Δt_{sk} is the time taken for the execution of the instruction. I_{sk} can be considered as a sampling from $\{I_j\}$ determined by the actual execution. Based on the above formulation, three types of programs, with no parallelism (sequential), with instruction parallelism and with input data parallelism (domain partitioning) are characterized below along with an illustration example.

Sequential program : A sequential program is characterized by the time dependencies between the execution instructions. Specifically, no two instructions overlap within their execution time intervals. Also, a specific ordering of instructions is followed in the execution. The conditions are expressed as

$$t_{0s_{(i+1)}} < t_{0s_i} < t_{0s_{(i-1)}} \text{ for } i = 1, \ldots, K \qquad (7.10.1)$$

and

$$\bigcap_{i=1}^{K} \tau_{s_i} = \emptyset. \qquad (7.10.2)$$

Let us consider an algorithm that involves an iterative improvements, similar to the *ME* algorithm for which the parallelism is attempted in this work. The structure of the program is written (iterations are denoted as time steps) as follows:

```
do 10  t = 1, number of time steps
    do 20  j = 1, Y-imagesize
        do 30  i = 1, X-imagesize
                image(t,i,j) = function (image(t-1,i,j), parameters(t-1))
        30 continue
    20 continue

    do 200  j = 1, Y-imagesize
        do 300  i = 1, X-imagesize
                parameters (t) = function(image(t,i,j))
        300 continue
    200 continue
10  continue
```

Instruction parallelism : On the Myrias system, parallelism is best exploited by identifying instruction parallelism, i.e identical instruction sequences that act on independent data sets. Starting from a sequential program, the data dependencies for given statement sequences are analysed and independent sequences are determined. The program can then be restructured to a parallel form. DO loops are natural candidates for this approach. Starting from the general program expression given by (7.8), the execution programs can be written as

$$P_e^T|P \equiv \left\{ (I_{s_k})_{D_k \subseteq \mathbf{D}}^{\tau_{s_k} \in T} \right\} \text{ for } k = 1,\ldots,K. \qquad (7.11)$$

where $\tau_{sk} = (t_{0sk}, \Delta t_{sk})$. I_{sk} is given by

$$\Gamma_{s_k} \equiv \left\{ \bigcup_{m=1}^{M} (I_{s_{km}})_{D_{km} \subseteq \mathbf{D}}^{\tau_{s_{km}} \in T} \right\} \text{ for } k = 1,\ldots,K \text{ and } m = 1,\ldots,M. \qquad (7.11.1)$$

where $\tau_{skm} = (t_{0km}, \Delta t_{km})$. $\tau_{skm} = \tau_{sk}$ for m = 1, ... ,M provided there exists independent instructions obeying the conditions:

$$I_{s_{nm,i \in k}} \equiv I_{s_{nm,n \neq k}} \; \forall m \qquad (7.12.1)$$

Also

$$\bigcap_{k} D_{km} = \varnothing \;\; \forall m.(D_{km} \in \mathbf{D}). \qquad (7.12.2)$$

For the instructions satisfying the conditions (7.12.1) and (7.12.2), such as the DO loops, the time condition in (7.10) becomes

$$t_{0s_{(i+1)}} = t_{0s_i} = t_{0s_{(i-1)}} \text{ for } i = 1, \ldots, K \qquad (7.13.1)$$

with

$$\bigcap_{i=1}^{K} \tau_{s_{im}} \neq \varnothing \; \forall m. \qquad (7.13.2)$$

The conditions in (7.13.1) and (7.13.2) represent instruction parallelism. If the conditions are not met then M = 1 and the execution proceeds sequentially. Under instruction

parallelism, the coding of the program will be as follows:

```
do 10 t = 1, number of time steps
   pardo 20 j = 1, Y-imagesize
      do 30 i = 1, X-imagesize
            image(t,i,j) = function (image(t-1,i,j), parameters(t-1))
      30 continue
   20 continue

   pardo 200 j = 1, Y-imagesize
      do 300 i = 1, X-imagesize
            parameters (t) = function(image(t,i,j))
      300 continue
   200 continue
10 continue
```

Domain partitioning: In this approach, the input data domain, i.e., the image domain, is partitioned into several sub-domains, based on the number of processors available. Under the partitioning scheme, the image domain is represented as

$$\mathbf{D}_N \equiv \bigcup_{i,j=1}^{\frac{N}{\sqrt{p}}} \mathbf{d}_{ij} \qquad (7.14)$$

The program expression in (7.8) is written as

$$P|\mathbf{D} \equiv \left\{ I_j \in P_g | \mathbf{d}_{mn} \in \mathbf{D} \right\}_{\mathbf{D}} \text{ for } j = 1,\ldots,J \qquad (7.15)$$

where the term P_g refers to a general program. The analytical approach followed in the above sections are the same except that the domain $\mathbf{D}$ is replaced by smaller domains. In writing (7.15), a square image has been assumed. $\mathbf{d}_{mn}$ in (7.15) is the partitioned domain each of area (N^2 /p), where p is the number of processing elements. In domain partitioning, the independency of image pixels has been used to remove the requirement for independence of data sets implied by (7.12.2) in using instruction parallelism. The approach however introduces the problems interms of mapping global *computations* over

the partitions. Coding the program will illustrate the issues underlying this approach.

```
pardo 10  t = 1, number of time steps
   pardo 20  k = 1, number of domains
      do 30  j = 1, Y-domainsize
         do 40  i = 1, X-domainsize
            domain(t,k,i,j)=function(domain(t-1,k,i,j), parameters(t-1))
         40 continue
      30 continue

      do 300  j = 1, Y-domainsize
         do 400  i = 1, X-domainsize
            parameters (t) = function(domain(t,k,i,j))
         400 continue
      300 continue
      20  continue
   10  continue
```

As we see from the program, the computations over images have been deliberately replaced by the same computations over domains. The approach makes use of the fact that most of the image processing computations are of block-type with operations restricted to a small neighbourhood of pixels. Investigations are however necessary for a given application problem to explore the suitablity of mapping. Another significant feature of the domain partitioning approach is the independency of the domain operations over time or iterations (pipelining). The necessity of sychronization at the end of each iteration is eliminated (asynchronous over iterations) which is expected to yield a significant performance improvement in a multi-processor environment. Although attractive, as mentioned above, investigation is necessary to explore the implementation possibilities for specific application problems.

7.2.3. Performance and efficiency measurements

Efficiency aspects of an algorithm in general are inferred using an empirical (*a posteriori*) or theoretical (*a priori*) or a combined approach [92]. In the theoretical approach, the execution time, memory space etc., needed by an algorithm is determined mathematically as a function of the sizes of the instances considered. The empirical approach on the other hand, consists of programming the algorithm and trying on different

instances with the help of a specific computer. The actual statistics are collected during the execution of the algorithm and the performance measures are estimated from these figures.

The performance of a parallel algorithm is in general measured in terms of speed-up-ratio. Assuming that speeds of the p processors are the same, the speed up ratio is defined as

$$S(p) = \frac{t(1)}{t(p)} \leq p \qquad (7.16)$$

where t(p) is the execution time using p processors (p>1) and t(1) is the execution time with a single processor. The quantity

$$\eta(p) = \frac{S(p)}{p} \qquad (7.11.1)$$

is termed as the efficiency of a parallel algorithm. In the speed-up and efficiency calculations, the execution time is often considered as the elapsed time of the process between the start of the program and finish of the last process. In an ideal parallel setup, a program should run in 1/p of the time required in a one processor setup. Graphically, an ideal case shows a linear (with slope = 1) relationship with the number of processors. In terms of the elapsed time statistics, the speed-up factor is calculated as

$$S(p) = \frac{e_t(1)}{e_t(p)} \qquad (7.16.1)$$

where e_t is the elapsed cpu time. The elapsed time is a function of the user, system, idle and wait times. For the definition and usage of these time figures in the Myrias parallel computer, the readers are referred to Appendix - G. The performance studies conducted using the Myrias system for image convolution and deconvolution algorithms are explained in the following section.

7.3. IMAGE CONVOLUTION ALGORITHM

Convolution and correlation are important spatial domain operations, central to

numerous image processing applications. Efforts are in constant progress to develop efficient algorithms for these basic operations [93], [76]. In the present context, the convolution operation is a core computation for the *ME* image deconvolution algorithm to be explained shortly.

7.3.1. Image convolution operation

The image convolution operation is defined by the relation

$$g(p,q) = \sum_{i=1}^{M} \sum_{j=1}^{M} h(i,j)\, f(p-i, q-j)) \text{ for } p = 1, \ldots, N \text{ and } q = 1, \ldots, N. \qquad (7.17)$$

h is a window of size M x M. The image is of dimension N x N. The coefficients of **h**, in general depend on the application. For example, if **f** is to be edge detected, it is convolved using (7.17) with 3 x 3 windows such as Laplacian, Prewitt and Sobel operators [76]. In the present study, the purpose of convolution is to blur the image **f** using a window filter **h**, say for example, the 7 x 7 worst case filter encountered in the earlier test studies of chapter 4. Graphically, convolution is an operation of moving the window **h** over the image **f**. Its pictorial illustration is shown in Fig.7.1. For each move, the convolution result is computed using (7.17). Wraparound errors are characteristics of discrete convolution, which appears as distortion around the end border of an image. This inevitable error near the image border is not objectionable. Periodic and mirror extension procedures may be used if necessary. The time complexity of the convolution algorithm is $O(M^2N^2)$.

7.3.2. Task specification strategy in the instruction parallelism approach

In general, the linearity, spatial invariance of **h**, and the localized window operation are the main features that are exploited for parallel implementation of the image convolution operation in (7.17). It is a special case, where the requirements (7.12.1) and (7.12.2) are

eliminated in the computational problem itself. Therefore the concepts of instruction and input domain partitioning are the same for image convolution. To implement (7.17) using instruction parallelism, on the SPS-2 system, *pardo* tasks are specified such that

$$T = \bigcup_{i=1}^{p} T_{ci} \tag{7.18}$$

where

$$T_{ci} = \bigcup_{\substack{j = \frac{iN}{p} + 1 \\ k = 1}}^{\substack{j = \frac{N}{p}(i+1) \\ k = N}} T_{cjk} \mid D_{jk} \tag{7.18.1}$$

The suffix c stands for a child task. Fig 7.2(a) shows the algorithm. As noted earlier (7.17) is a special case with the task domain D equal to the input domain $\mathbf{D}$. The stored locations are independent for each child and also the tasks are of duplicated type and hence the merging process is quite straightforward.

7.3.3. Performance studies with instruction parallelism

The algorithm was run on a simple computer generated image (checker board pattern) of size 256 x 256 shown in Fig. 7.3. The speed-up-ratios were calculated from the elapsed time statistics using (7.16.1). They are plotted in Fig 7.4. The ideal performance curves are also shown for comparison. The elapsed time plot shows the time improvements very close to the ideal. The measurements indicate a 91% efficiency with 16 processors.

The speed-up characteristics shows unexpected jumps at the points marked as p_1 and p_2 in Fig. 7.4, the reasons for which are investigated below. As seen from the task specification strategy in (7.18), the child tasks are determined by the ratio N/p. If the tasks

are not exactly divided by the ratio, the tasks become uneven, generating an uneven load of the allocated processors. The performance of the algorithm is then influenced by the distribution of the tasks. With respect to (7.18.1)

$$T_{ci} = \bigcup_{\substack{j = \frac{iN}{p} + 1 \\ k = 1}}^{\substack{k = N \\ j = \left(\frac{N}{p}(i+1)\right) + n_a(1-\gamma)}} T_{cjk} \mid D_{jk} \qquad (7.19)$$

γ in (7.19) is a 'load distribution' parameter, with its value 1 for some processors and 0 for others. The processors with $\gamma = 0$, face an additional amount of work, (in the present context it is processing of an additional n_a columns of data), during which time, the processors with $\gamma = 1$ remain idle.

This situation can be explained in practical terms from the recorded elapsed times shown in Appendix - H. in the execution of the convolution algorithm. Let us consider point 9, which corresponds to a resource demand of 50 processors. In the algorithm, 50 parallel tasks would have been generated, one for each processor. For 256 tasks, some processors have a task space of 5 x 256 and some 6 x 256. Let us say that 44 of the tasks do calculations on 5 colums and 6 tasks do 6 columns each. After the shorter tasks have finished, the six longer ones have one more column to process upon. This time (average) can be calculated from the number of tasks and the user time statistics. In this case it is 0.3098 secs/column. When the longer tasks are being executed, the shorter tasks, that have finished their jobs will be idle for 44 x 0.3 = 13.2 seconds. This constitutes 62.6% of the recorded idle time. This accounts for a reduction of 76.3% in the elapsed time. Therefore, the sudden decrease at point 8 on the speed-up characteristics is the result of the sudden increase in the idle time. The situation is similar between points 9 and 10 because point 10 pertains to a processor resource of 60. At point 11, the tasks get divided evenly among the

processors. Hence there are no idle time discrepancies due to shorter and and longer tasks. The idle time behaviour returns to its 'normal'.

7.3.4. Performance studies using domain partitioning

For the domain partitioning approach, the big (square) image domain is divided into square sub-domains d_{jk} (j = 1, ..., N/√p and k = 1,.....,N/√p). By this partitioning scheme, the tasks are specified at the input data level. Mathematically

$$T = \bigcup_{i=1}^{p} T_{ci} \tag{7.20}$$

where

$$T_{ci} = \bigcup_{\substack{j=1 \\ k=1}}^{\substack{k=\frac{N}{\sqrt{p}} \\ j=\frac{N}{\sqrt{p}}}} T \mid d_{jk}. \tag{7.20.1}$$

Each domain is 'seen' as an individual image by each processor in this case. In this mode, as indicated earlier, some considerations are necessary to ensure the functional correctness of the resulting convolution operation. The domains need to communicate with their neighbouring domains for reading their input data so as to eliminate the wrap around errors within its own 'private' domain. One way of resolving this problem would be to increase the visibility of each domain, with the task domain restricted to the private domain. In terms of the expressions, (7.20) remains the same, but the partitioned domain area is increased from N^2 to (N+M-1)(N+M-1). The algorithm is shown in Fig. 7.2(b).

During execution, when a task references an address in the visible task domain, but

the page which stores that value is not present, the operating system sends a copy of that page to the task. The pages modified by the tasks are not sent back to the parent task until all the child tasks have completed. So there is no communication between sibling tasks. However, the overhead involved in the process of implicit communication affects the overall time performance of the algorithm. The test results showed a 27% increase (with 16 processors) in the elapsed time compared to the instruction parallelism approach, where the communication requirements are relatively minimum. The performance efficiency in the domain partitioning approach was found to be 88.5% which is almost equal to that of the instruction parallelism approach. The results show that the SPS system is not well-suited to the domain partitioning approach of convolution mainly because of the system overhead in meeting the communication requirements. The result of convolution was observed to be the same in both cases and is shown in Fig. 7.5. It is used as an input image for the *ME* deconvolution algorithm discussed below.

7.4. IMAGE DECONVOLUTION-THE *ME* ALGORITHM

Starting from the image in Fig. 7.5, the purpose of image deconvolution is to restore the source image shown in Fig. 7.3 using the *ME* method. The solution is obtained using an algorithm based on the constrained parameter relaxation approach explained under section 7.1.3. Here, its implementation is attempted using the SPS system. The various computational stages of the algorithm were already explained in chapter 4 (see Fig. 4.1). Let us consider the data processing phase of the algorithm, which consists of the data space (phase-2) and the parameter space (phase-3) calculations. The parameter space calculations have a time complexity of $O(N^2)$. The calculation of $\partial Q/\partial f$ involves two successive convolution operations with a time complexity of $O(N^2M^2)$. Both measures exclude the influence of the number of iterations over the problem size. This, as explained in chapters 3 and 4, is due to the procedure having a mechanism to adjust the number of iterations. This flexibility allows the assumption of a constant time behaviour of the algorithm within the

limits of the mechanism. Therefore, for all practical purposes, the time complexity of the *ME* deconvolution algorithm is the same as the convolution operation. The computational involvement of each phase in the algorithm is calculated as a percentage of their elapsed cpu times in the 1-processor set-up, and is shown as a pie-chart in Fig. 7.6. The figure shows that the calculation of $\partial Q/\partial f$ amounts to 88% of the total computational time.

7.4.1. Instruction parallelism approach

As in convolution, the different phases of the algorithm in a sequential program was restructured for instruction parallelism and is shown in Fig.7.7, highlighting the estimate and data space calculations. In executing the parallel algorithm, the elapsed time characteristic of each phase was recorded. Since our main concern is the computational part of the algorithm, the phase-2 and 3 of the algorithm are studied for their performance in detail. The algorithm was run on problem sizes of 64 x 64 and 256 x 256.

The calculation of $\partial Q/\partial f$, as mentioned above, consists of two spatial convolution operations. The program steps for this phase are the same as that of the convolution algorithm explained in section 7.3. The recorded performance of the algorithm for this phase is shown in Fig. 7.8. The reduced linearity in the speed-up is due to the reduced parallelism in the algorithm. This is because, the two convolution operations remain sequential in the program and could not be executed in parallel efficiently using instruction parallelism. The parallel execution of phase-3 which involves parameter and estimate calculations makes use of the fact that the parameters λ and β are constant multiplying factors for all the pixels in the procedure. The speed-up characteristics of this phase is shown in Fig. 7.8 along with the resulting over all speed - up characteristics in Fig. 7.9.

From the recorded speed-up characteristics, it is possible to obtain an optimum number of processors for a satisfactory performance. The criterion for optimum performance is to have maximum possible efficiency of the algorithm for the minimum possible elapsed time. The overall efficiency and the elapsed time (average time for one

iteration) variations are shown in Fig. 7.10, for both 64 x 64 and 256 x 256 problem sizes. Graphically, the optimum corresponds to the point of maximum separation distance of the two curves. In the case of 64 x 64 image, the optimum number of processors corresponding to these figures is 4. It gives 74.1% efficiency with 50% increase in the computational speed. For an optimum number of 8 processors in the 256 x 256 case, the efficiency of the algorithm is 78% with the computational speed improvement of nearly 99%. The 256 x256 problem size shows a better performance compared with the 64 x 64 case. This is because of the increased work for each child in the case of 256 x 256 problem sizes.

The result of the algorithm is shown in Fig. 7.11. Visual comparison of the result with the source image in Fig. 7.3 shows a satisfactory functional performance of the algorithm. The behaviour characteristics of the algorithm are shown in Fig.7.12. The variations are consistent with the predicted characteristics discussed in chapter 4. The chosen value of K_λ was 100 for the 256 x 256 problem size, and convergence was obtained at iteration 14. As discussed earlier, K_λ can be adjusted for faster convergence.

7.4.2. Domain partitioning approach - a discussion

The implementation of the algorithm in the domain partitioning approach follows the same relation given by (7.19) with the task T corresponding to the *ME* restoration algorithm. As discussed earlier, the degree of concurrency is higher in the approach, but the communication requirements to realize the functional correctness of $\partial Q/\partial f$ need significant attention. Following the visibility extension for functional correctness, the domain area $\mathbf{d}_{jk}$ for the first convolution first convolution increases to $(N/p + 2(M-1))^2$. The active child task domain becomes

$$k = \frac{N}{\sqrt{p}} + M\text{-}1$$

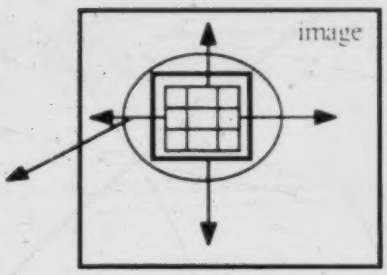
$$j = \frac{N}{\sqrt{p}} + M\text{-}1$$

$$T_{ci} = \bigcup_{\substack{j=1 \\ k=1}} T\left[d_{jk}\right]. \qquad (7.17)$$

For the execution of the second convolution, the visible and the active task domain size becomes the same as that of a single convolution case explained in 7.3.4. The extension of domain areas consitute an overlapped domain configuration. The studies on the activities within the overlap regions in the light of synchronous and asynchronous requirements, the implementation of the different overlap schemes using Myrias computers are in progress and is not elaborated here.

The discussions show that the parallell algorithms in general offer significant time improvements over the sequential algorithms. In image convolution for 256 x 256 image, the time improvement was as high as 98% and in image deconvolution it was 89% over the sequential algorithms. In this study, it was found that the instruction parallelism was found suitable for a local and sychronous linear operations such as the image convolution with an efficiency figure as high as 98%. In a typical nonlinear environment, such as *ME* deconvolution algorithm, the performance slows down and the flexibility to increase parallelism in the program is found to be minimal. The input data parallelism basis of the domain partitioning approach, on the other hand offers an effective solution to increase parallelism. A study in this direction was however not easily feasible because of the mapping problems and the increased system overhead associated with their implementation using the SPS-system.

h(-1,-1) f(x-1,y-1)	h(-1,0) f(x-1,y)	h(-1,1) f(x,y+1)
h(0,-1) f(x,y-1)	h(0,0) f(x,y)	h(0,1) f(x,y+1)
h(1,-1) f(x+1,y-1)	h(1,0) f(x+1,y)	h(1,1) f(x+1,y+1)



$$g(x,y) = h(-1,-1)f(x-1,y-1) + h(-1,0)f(x-1,y)+h(-1,1)f(x,y+1) + h(0,-1)f(x,y-1)+h(0,0)f(x,y)+h(0,1)f(x,y+1)+ h(1,-1)f(x+1,y-1)+h(1,0)f(x+1,y)+h(1,1)f(x+1,y+1)$$

Fig. 7.1 Image convolution operation

```
DO BEGIN
IN PARALLEL FOR  1 ≤ j ≤ number of
                 processors
 find start(j) and end (j)
 DO 10  start(j) ≤ i ≤ end(j)
  DO 20  1 ≤ k ≤ N
   calculate g(k,i)
20 CONTINUE
10 CONTINUE
END IN PARALLEL DO
```

(a)

```
DO BEGIN
 IN PARALLEL FOR  1 ≤ j ≤ number of
                  processors
  find startx(j) and endx(j)
  find starty(j) and endy(j)
 DO 10 startx(j) ≤ i ≤ endx(j)
 DO 20 starty(j) ≤ k ≤ end y(j)
   calculate g(k,i)
20  CONTINUE
10  CONTINUE
 END IN PARALLEL DO
```

(b)

Fig.7.2 Image convolution using instruction and data parallelism
(a) instruction parallelism (b) data parallelism

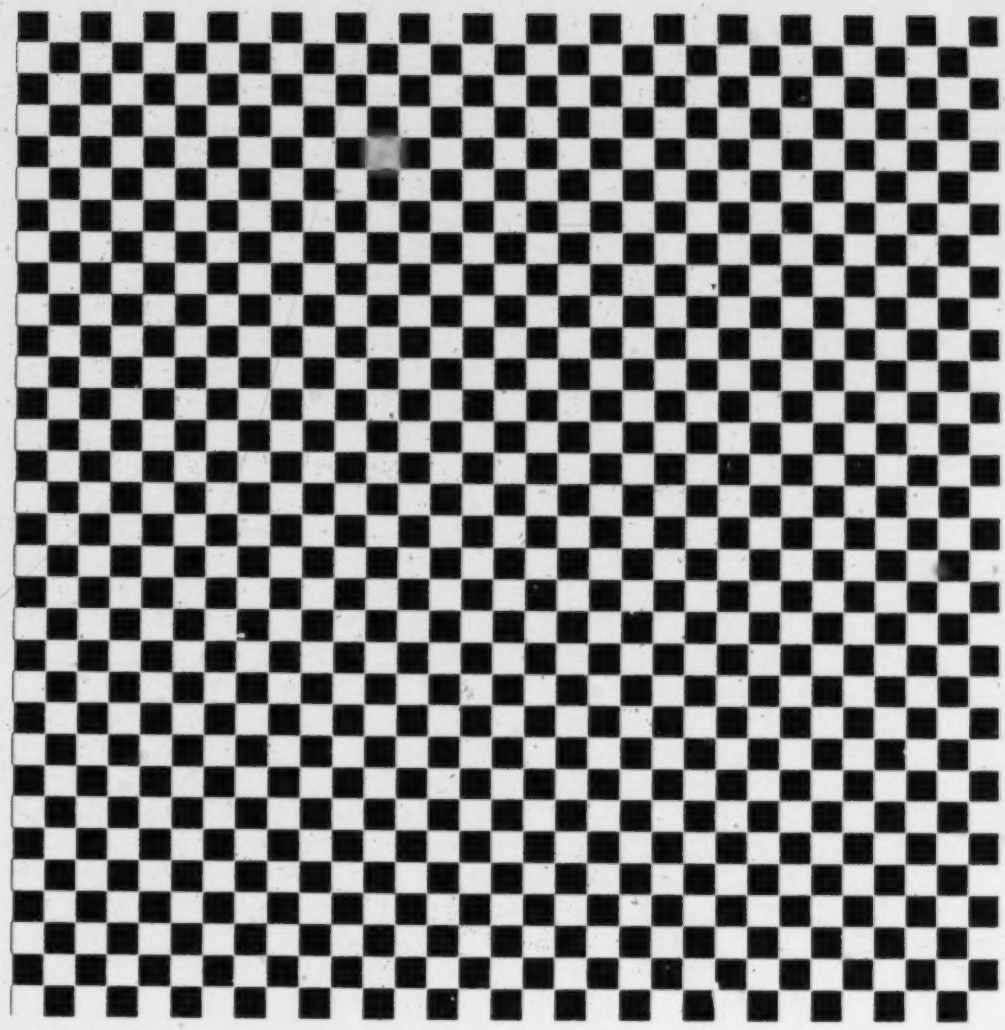
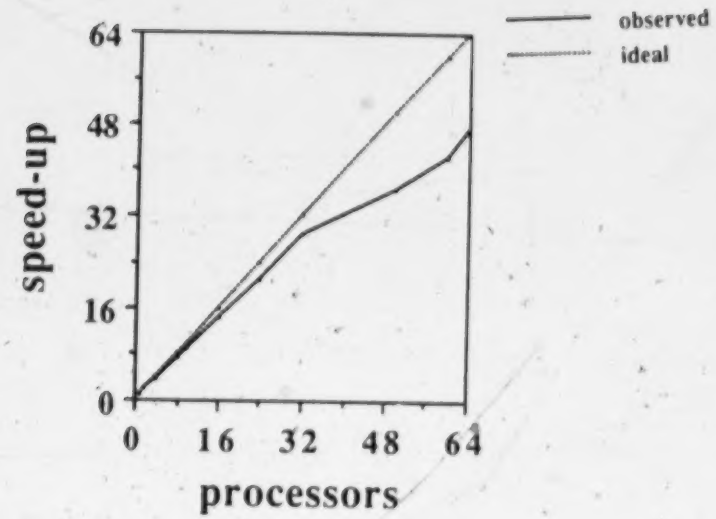


Fig.7.3 Input image for the convolution algorithm



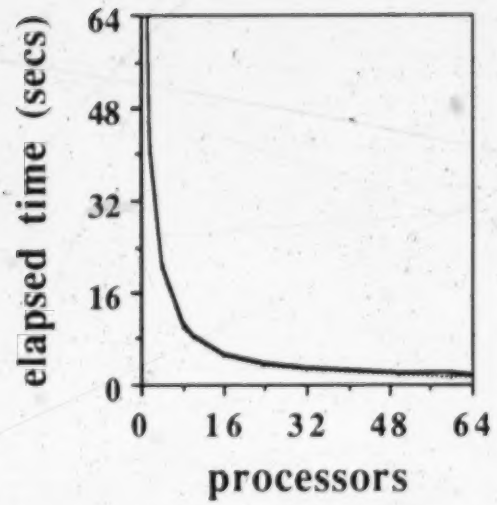


Fig.7.4 Speed - up and elapsed time performance characteristics

Fig.7.5 Output of the convolution algorithm

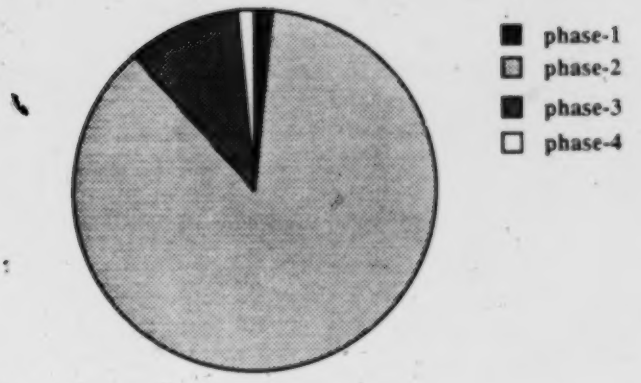


Fig.7.6 Time involvement of different computational phases in *ME* deconvolution

```
      i = 1
      start : estimate(0)    and input d
100   IN PARALLEL
      convolve  estimate(i - 1)
100   END PARALLEL CONVOLUTION
      calculate   Q
200   IN PARALLEL
      convolve  for ∂q/∂f calculation
200   END PARALLEL CONVOLUTION
      calculate   λ  and  β
300   I N PARALLEL
      calculate the   ME  estimate   and q
300   END IN PARALLEL
      i = i+1   ; GOTO 100
```

data space calculations (phase - 2)

estimate space calculations (phase - 3)

Fig.7.7 Image deconvolution algorithm using instruction parallelism

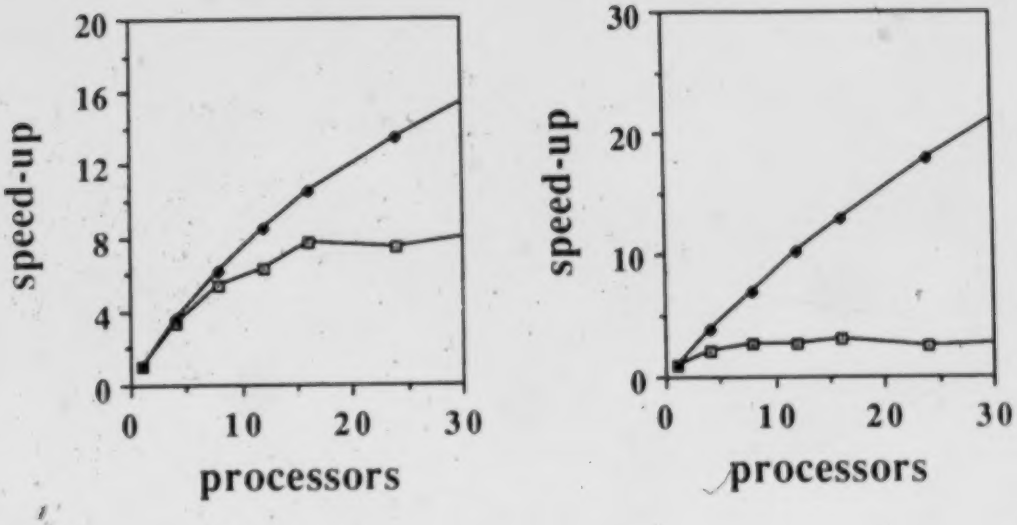


Fig.7.8 Speed - up characteristics for data space and estimate space calculations

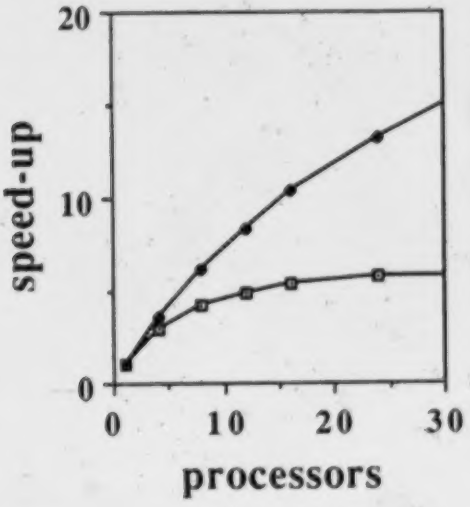
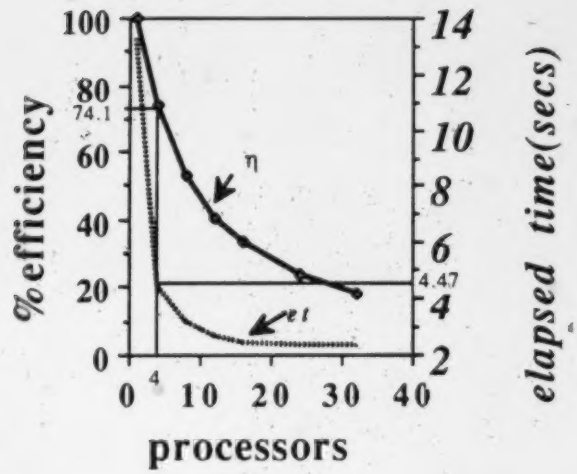


Fig.7.9 Overall speed - up characteristics of the algorithm



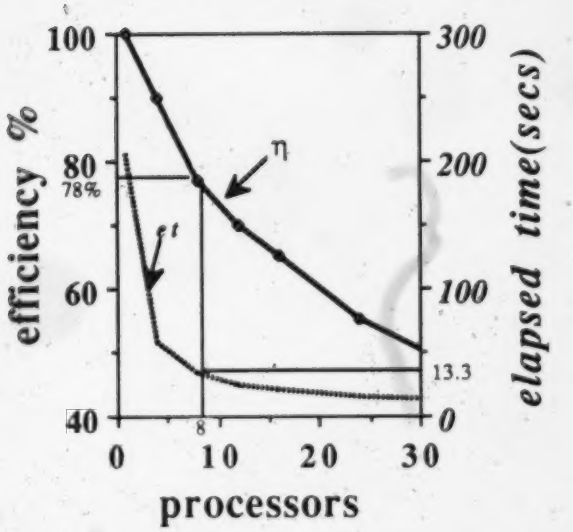


Fig.7.10 Estimation of optimum number of processors

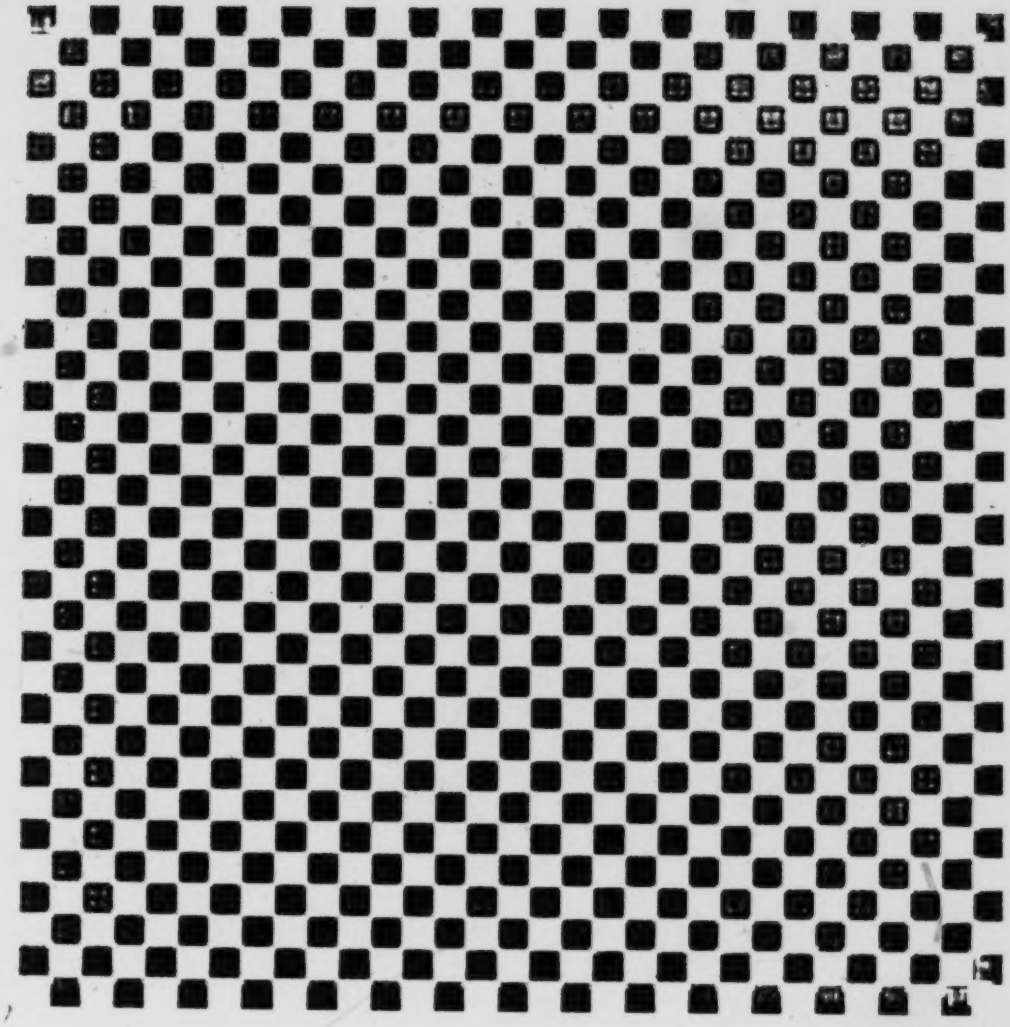


Fig.7.11 Output of the image deconvolution algorithm

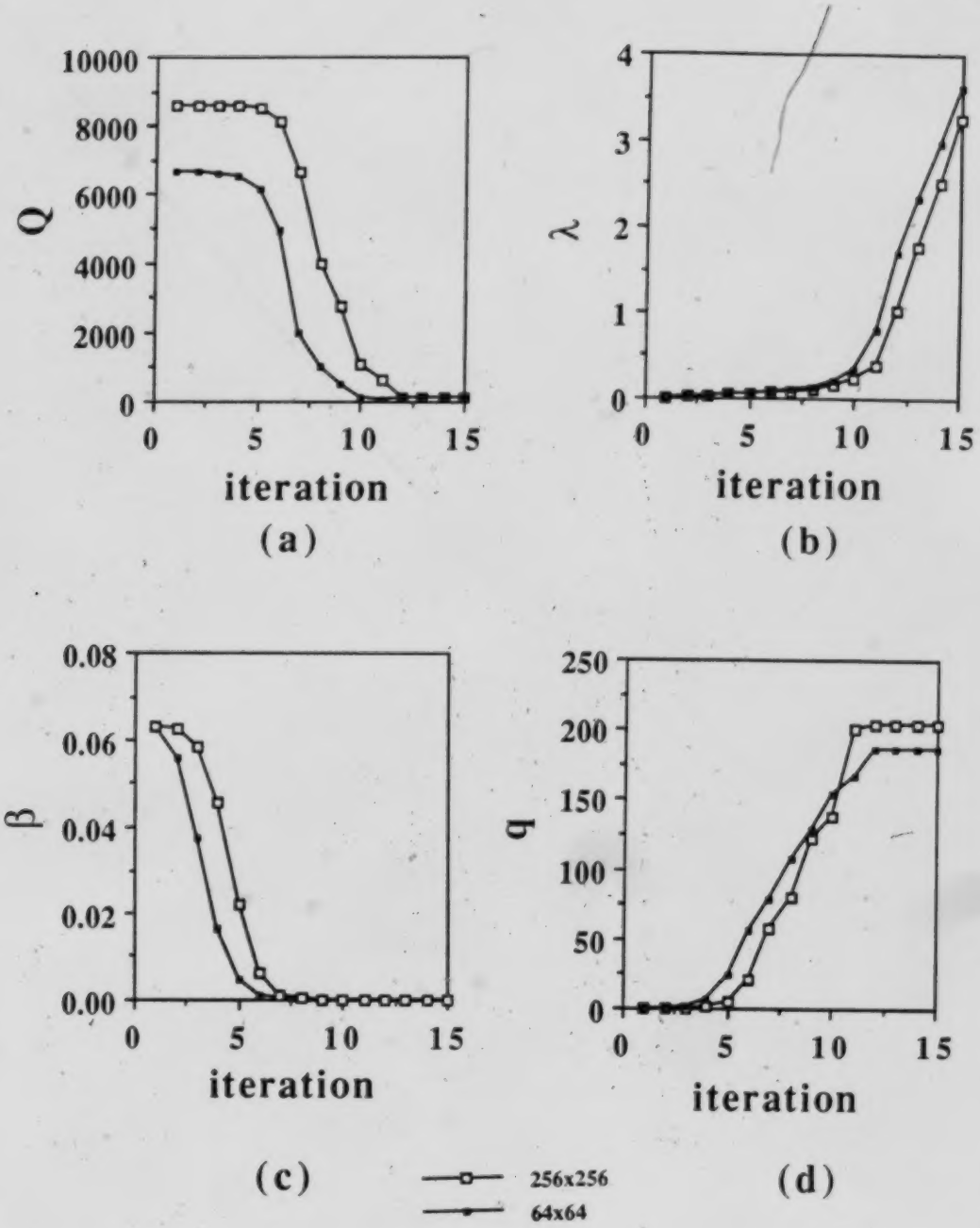


Fig.7.12 Behaviour characteristics of the relaxation procedure in a parallel algorithm
Variations of (a) Q; (b) λ; (c) β and (d) q.

CHAPTER 8

CONCLUSIONS

The primary aim of this thesis is to understand and contribute to the progress of the maximum entropy (*ME*) method in the areas of image processing and machine vision. In reference to the aim, the status and the foundations of the method were investigated in chapter 1 and 2 as applied to images, followed by a series of developments in the subsequent chapters. Conclusions are summarized below.

8.1. The *ME* method-a general outlook

From the discussions of chapter 1 and 2, it is understood that the *ME* method stems from two different concepts: physical and information theoretic. To summarize, the problem of estimating **f** from **d** in (2.2) is understood by the 'physical' group as the problem of restoring a true image based on the statistical considerations of image formation, with prior knowledge and the entropy form determined by the type of application problem. The information theoretic group, on the other hand, believes that finding the true image is an inference problem of finding a distribution **f** that incorporates the testable information in **d** as constraints with a suitable prior information. Both concepts make use of the entropy functional form $plog(p/q)$ with q as the probability distribution resulting from prior information and p resulting from the testable information **d**. In a more broad sense, the method is expected to be useful in application areas, ranging from the traditional inverse problems to modern artificial intelligence problems of machine reasoning, learning and problem solving. But investigations show that the progress of the *ME* method is limited to inverse problems. Within the class of inverse problems, the method is not firmly established beyond 1-D power spectrum estimation. This thesis has adopted the viewpoint that the present status of the method requires application

developments for further progress. The issues of importance for application developments are: the general ways of estimating prior probability distributions and a general basis for obtaining *ME* solution. In the literature, specifically for image processing applications, as discussed in chapter 2, the computational efforts are minimal compared to the theoretical developments. Considering the importance of a general computational scheme for the present as well as the future applications, attention is focused in this thesis on the computational developments. Application studies have led to interesting observations and conclusions for the *ME* method for further studies.

8.2. Computational scheme for the *ME* method

A parameter controlled relaxation approach is proposed for the *ME* method in this thesis. In this scheme, an initial estimate undergoes successive transitions towards the final solution. The transitions are best explained in terms of states of a system. The initial state of the system is determined by the prior knowledge conditions. In this thesis work, a uniform state results from the uniform prior probability estimate but in general, is not a strict requirement for the solution scheme. the successive states of the system are determined by the stationary point equation of the Bayesian based *ME* objective function. The state transitions are controlled by a set of constraints and parameters: Intermediate constraints and Lagrange parameter(s) λ for consistency with the input information constraint(s); relaxation parameter(s) β to adaptively control the influence of the external constraints upon the system for improved numerical stability and convergence of the system; upper-bound constraint for the required input-scaling. The behaviour of the system resulting from these controls is characterized in terms of the magnitude of the logarithmic state entropy and its average. The states with entropy maximum is the required state, converged for the given input constraints. The use of function Q in the Bayesian formulation of the *ME* method explained under section 2.8.2 reduces the size of the control parameters from $\mathbf{m}$ (the total number of pixels in an image) to 1. The above inferences and

the test studies reported in this thesis are based on the Bayesian formulation of the *ME* method using the function Q. It is however anticipated that the same inferences can be extended to the standard case of **m**-Lagrange mutipliers.

8.3. Application studies in Bayesian restoration

The algorithm based on the relaxation scheme with the Bayesian fomulation of the *ME* method was applied to different types of degraded images for restoration. In the case of inputs degraded by Gaussian noise, analysis of the error behaviour of the algorithm shows that the specification of noise variance is not necessary. The numerical behaviour of the algorithm was found consistent with the analytical results, giving rise to a stable and convergent solution for varying types of inputs. The rate of convergence is user-controlled in the algorithm by a component K_λ of the parameter λ. No restriction is observed for the possible range values of K_λ with respect to the numerical behaviour of the algorithm. The restoration performance of the algorithm is also found consistent with minimum artifacts even under the worst conditions of degradation. With the worst case examples, a mean square error evaluation of the *ME* restoration performance shows an improvement of 58% and 87% over the constrained least squares approach. Under minimal degradation conditions, the algorithm shows unexpected enhancement effects in the restored results with the effects becoming increasingly pronounced upon and after convergence. Analysis reveals a Laplacian operation underlying the enhancement effects in the *ME* restoration performance. The increased sensitivity of the Laplacian mechanism to the minimum degradation test conditions are shown to be the reason for the prominance of enhancement effects upon convergence.

One of the nontrivial requirements of an algorithm in application studies is its capability to deal with realistic images, for which the degradation parameters are not known. Even if the degradation parameters are determined experimentally, the sensitivity of

the algorithm performance to the accuracy of the degradation parameters is an important issue for practical and real time applications. In this context, the elimination of Gaussian noise variance specifications proposed in this thesis studies, partly contributes to over come the experimental difficulties. However, as demonstrated in chapter 4, the influence of the level of noise on the noise reduction capabilities of the algorithm leads to the problem of settling for an optimum level of restoration performance. Development of independent control schemes for noise reduction and schemes for 'blind deconvolution' (elimination of the point spread function specification) are significant practical issues of future interests for the *ME* relaxation algorithm.

8.4. Links with Biological visual models

Test results of the *ME* restoration algorithm under minumum degradation conditions have also led to new concepts in this thesis study. With the help of specially devised test schemes for no-degradation conditions, the *ME* relaxation algorithm is shown to simulate the psycho-physical characteristics of Mach bands of biological visual systems for simple intensity gradient stimulus patterns. Marked similarities are observed, particularly the asymmetries and the response modulation with slope variations of the stimulus pattern. A suitable parameter setting is found to be the key feature for the algorithm in forming the bands. Analytical results show a Laplacian mechanism for the parameter settings. The reasons for the modulation behaviour with slope variations are not clear yet. The response of the algorithm with complex Mach patterns are not found favourable. The difficulties of determining a suitable parameter settings for the complex patterns are believed to be the reason for the unfavourable responses. It is anticipated that the relaxation algorithm based on the $\mathbf{m}$-Lagrange multiplier *ME* method has a greater potential for more interesting simulation studies of biological visual responses.

In another analytical study, it is found that the relaxation scheme shows similarities to Mach's biological visual model but with one major difference. Unlike Mach's model,

which is non-linear stimulus dependent, *ME* relaxtion algorithm is non-linear response dependent. This difference accounts for better results possible with the *ME* algorithm compared to Mach's model with respect to the simulation of Mach band characteristics.

Test studies also suggest new application prospects in edge detection and enhancement for the *ME* method. The application differs from the current approaches in the sense that *ME* edge enhancement operation does not require external Laplacian masks or spatial derivative masks but may require suitable functional derivative operators, a study which needs further investigations for stable inferences.

8.5. Parallel implementation studies

Implementation of relaxation algorithms, in general have a favourable status in current VLSI computer technology. With an aim of implementing a real-time *ME* restoration algorithm using dedicated parallel architectures, the concepts of Instruction parallelism and the domain partitioning parallelism were explained. Initial studies have been performed on Myrias parallel computer that has an MIMD architecture. Studies made on specific image convolution and *ME* image deconvolution algorithms using instruction parallelism and domain partitioning techniques lead to the following conclusions: Image convolution, being a linear and spatial invariant operation, fulfills the requirements of the instruction parallelism, leading to an efficiency of 91% with 16 processors. On the other hand, with the *ME* deconvolution algorithm with strong data dependencies, the instruction parallelism approach was not found satisfactory. For 256 x 256 images, studies show 78% efficiency and 99% time improvement (over a single processor) with an optimum of 8 processors. The understanding of the domain partitioning techniques shows a potential for significantly improved performance, but stable results could not be obtained due to the system overhead in its SPS implementation. The possible schemes for partitioning the computations of *ME* algorithm are being worked out.

BIBLIOGRAPHY

[1] E. T. Jaynes, "Where Do We Stand on Maximum Entropy?", in *The Maximum Entropy Formalism*, editors: R. D. Levine and M. Tribus, *The MIT Press*, Massachussets (1979) 15-118.

[2] E. T. Jaynes, "Probability Theory in Science and Engineering ", *colloquium lectures in pure and applied science*, No. 4, Socony Mobil Oil Company, Inc. U.S.A, (1959) 1-21.

[3] C. E. Shannon, "A Mathematical Theory of Communication", *Bell System Tech. Journal*, No., 27, (1948) 379-423 and 623-656.

[4] E. T. Jaynes, " Information Theory and Statistical Mechanics I", *Phys. Rev.*, Vol. 106, (1957) 620-630.

[5] E. T. Jaynes, "Information Theory and Statistical Mechanics II", *Phys. Rev.*, Vol. 108, (1957) 171-190.

[6] S. Kullback, Information Theory and Statistics. *Newyork: Wiley*, 1959.

[7] J. P. Burg, "Maximum Entropy Spectral Analysis" in *Proc. 37th meeting. Society of Exploration Geophysicists*, Oklahoma city, OK, Oct.31, (1967).

[8] E. A. Robinson, "A Historical Perspective of Spectrum Estimation", *Proc. of IEEE*, Vol. 70, No. 9, (1982) 885-907.

[9] A. Papoulis, "Maximum Entropy and Spectral Estimation: A Review", *IEEE Trans., Acoust., Speech., Signal Processing*, Vol. ASSP-29, (1981) 1176-1186.

[10] T. J. Ulrych, "Spectral Analysis and Time Series Models: A Geophysical Perspective" in Maximum-Entropy and Bayesian Methods in Inverse Problems, editors: C. R. Smith and W. T. Grandy. Jr., *D. Reidel Publishing Company, Dordrecht*, Holland (1985) 243-271.

[11] B. Roy Frieden, "Restoring with Maximum Likelihood and Maximum Entropy", *Journal of the optical society of America*, Vol 62, No. 4 (1972)511-518.

[12] S. J. Wernecke and L. R. D'Addario, "Maximum Entropy Image Reconstruction", *IEEE Trans. Comp*, c-26, (1977) 351-364.

[13] S. F. Gull and G. J. Daniell, "Image Reconstruction from Incomplete and Noisy Data", *Nature*, Vol. 272 (1978) 687-690.

[14] R. Willingale, "Use of the Maximum Entropy Method in X-ray Astronomy", *Mon. Not. R. astr. Soc.*, 194, (1981)359-364.

[15] R. K. Bryan and J. Skilling, "Deconvolution by Maximum entropy, as illustrated by application to the jet of M87", *Mon. Not. R. astr. Soc.*, 191, (1980) 69-79.

[16] S. F. Gull and J. Skilling, "Maximum Entropy Method in Image Processing", *IEE Proc.*, Vol .131, Pt. F, No. 6, (1984) 646-659.

[17] G. J. Daniell and S. F. Gull, "Maximum Entropy Algorithm Applied to Image Enhancement ", *IEE Proc.*, Vol.127, pt. E, No. 5. (1980) 170-172.

[18] S. F. Burch , S. F. Gull and J. Skilling, "Image Restoration by a powerful Maximum entropy method", *Computer vision, Graphics and Image processing*, Vol. 23 (1983) 113-128.

[19] B. R. Frieden, "Maximum Entropy Restorations of Ganymede", *SPIE, Opt. Soc. Am.*, 74, (1976) 160-165.

[20] A. Mohammad-Djafari and G. Demoment, "Maximum Entropy Image Reconstruction in X-Ray and Diffraction Tomography", *IEEE Trans. Med. Imaging*, vol. 7, No. 4, (1988) 345-354.

[21] J. Skilling and S. F. Gull, "Algorithms and Applications" in *Maximum-Entropy and Bayesian Methods in Inverse Problems*, editors: C. R. Smith and W. T. Grandy. Jr., D. Reidel Publishing Company, Dordrecht, Holland (1985) 83-132.

[22] J. Skilling, "The Axioms of Maximum Entropy", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: Gary J. Erickson, C. R. Smith, Vol. 2: Applications, Kluwer academic publishers, Dordrecht, Boston, London (1988) 173-187.

[23] A. K. Livesey and J. Skilling, "Maximum Entropy Theory", *Acta. Cryst.*, A41, (1985) 113-122.

[24] S. F. Gull and T. J. Newton, "Maximum Entropy Tomography", *Applied Optics*, Vol. 25, No. 1 (1986) 156-160.

[25] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", in *Readings in Computer Vision: Issues, Problems, principles and Paradigms*, editors: M. A. Fischler and O. Firschein,

Morgan Kaufmann Publishers, Inc., Los Altos, CA 94022 (1987) 562-584.

[26] A. Lippman, "A Maximum Entropy Method for Expert System Construction" in *Maximum-Entropy and Bayesian Methods in Science and Engineering, Vol.2: Applications,* editors: Gary J. Erickson, C. R. Smith, Kluwer academic publishers, Dordrecht, Boston, London (1988) 243-263.

[27] S. Geman, " Stochastic Relaxation Methods for Image restoration and Expert systems", *Maximum-Entropy and Bayesian Methods in Science and Engineering, Vol.2: Applications,* editors: Gary J. Erickson, C. R. Smith, Kluwer academic publishers, Dordrecht, Boston, London (1988) 265-311.

[28] W. K. Pratt, " Digital Image Processing", John wiley and sons, New york (1978)

[29] E. L. Hall, "Computer Image Processing and Recognition", Academic Press, Inc., New York 10003 (1979).

[30] R. C. Gonzalez and P. Wintz, "Digital Image Processing", Addison-wesley publishing company, MA, (1977).

[31] A. Rosenfeld, "Image Analysis: Problems, Progress and Prospects", *Readings in Computer Vision: Issues, Problems, principles and Paradigms,* editors: M. A. Fischler and O. Firschein, Morgan Kaufmann Publishers, Inc., Los Altos, CA 94022 (1987) 3-12.

[32] M. D. Levine, "Vision in Man and Machine", McGraw-Hill, New York (1985).

[33] M. Bertero, T. A. Poggio and V. Torre, "Ill-Posed Problems in Early Vision", *Proc. of IEEE.,* Vol. 76, No. 8 (1988) 869-888.

[34] T. Poggio, V. Torre and C. Koch, "Computational vision and Regularization Thoery", *Readings in Computer Vision: Issues, Problems, Principles and Paradigms*, editors: M. A. Fischler and O. Firschein, Morgan Kaufmann Publishers, Inc., Los Altos, CA 94022 (1987) 3-12.

[35] O. Kuebler and M. Kunt, " Modeling, analysing and Coding images: A tutorial Digest", *Signal processing II: Theories and applications*, edited by. H.W. Schussler, Elsevier Science Publishers, Eurasip (1983) 119-126.

[36] W. T. Grandy, "Incomplete and Generalized Inverses ", *Maximum-Entropy and Bayesian Methods in Inverse Problems*, editors: C. R. Smith and W. T. Grandy. Jr., D. Reidel Publishing Company, Dordrecht, Holland (1985) 1-19.

[37] D. E. Dudgeon and R. M. Mersereau, "Inverse Problems", *Multidimensional Signal Processing*, Prentice-Hall signal processing series, NewJersey, (1984) 348-390.

[38] E. T. Jaynes, "Where do we go from here?", *Maximum-Entropy and Bayesian Methods in Inverse Problems*, editors: C. R. Smith and W. T. Grandy. Jr., *D.* Reidel Publishing Company, Dordrecht, Holland (1985) 21-58.

[39] H. C. Andrews and B. R. Hunt, "Digital Image Restoration", Prentice-Hall, England (1977).

[40] D. Knuth, "Fundamental Algorithms" Addition-Wesley Publishing Company, inc., (1973).

[41] G. Brassard and P. Bratley, "Algorithmics, Theory and Practice" Prentice-Hall, Englewood Cliffs, Newjersey, (1988).

[42] B. R. Hunt and O. Kuebler, "Karhunen-Loeve Multispectral Image Restoration", Part 1: Theory, *IEEE trans. Acoust., Speech., Signal Processing, vol. ASSP-32, No. 3,* (1984) 592-600.

[43] R. Kikuchi and B. H. Soffer, "Maximum Entropy Image Restoration. I. The entropy expression", *J. Opt. Soc. Am., vol. 67,* No. 12, (1977) 1656-1665.

[44] H. J. Trussell, "A Priori Knowledge in Algebraic Reconstruction", *Advances in Computer Vision and Image processing, A research annual, Vol. 1: Image Reconstruction from Incomplete Observations,* editor: T. S. Huang, Jai press inc., connecticut (1984) 265-316.

[45] E. T. Jaynes, " Prior Probabilities", *IEEE Trans. Sys. Sci. and Cyb,* Vol. SSC-4, No., 3, (1968) 227-241.

[46] E. T. Jaynes, "On the Rationale of Maximum Entropy Methods", *Proc. of IEEE,* Vol. 70, No. 9 (1982) 939-952.

[47] K. Hanson, "Bayesian and Related Methods in Image Reconstruction from Incomplete Data", *Image Recovery, Theory and Applications,* editor: H. Stark, Academic Press, Inc., San Diego, California (1987) 79-121.

[48] J. D. O' Sullivan and M. M. Komesaroff, "Maximum Entropy-Some Reservations", *indirect imaging, Proc., of Int., Symp.,* Cambridge University Press, Sept 1983, Sydney, Cambridge, (1984) 297-302.

[49] B. R. Frieden, " The Axiomatic approach", *Probaility, Statistical Optics, and Data Testing,* Springer-Verlag, Heidelberg, (1983) 7-36.

[50] S. I. Maximov. The Universal Maximum Entropy Reconstruction of Radio Brightness. *Radio Science*, Vol. 17, No. 5, (1982), 831–844.

[51] J. Skilling and S. F. Gull. The Entropy of an Image. *SIAM-AMS proceedings*, vol. 14, (1984), 167–189.

[52] S. F. Gull and J. Skilling. The Entropy of an Image. *Maximum Entropy and Bayesian Methods in Inverse Problems*, editors C. R. Smith and W. T. Grandy, Jr. (D. Reidel Publishing Company, Dordrecht, Holland), (1985), 287–301.

[53] B. R. Frieden. Restoring with Maximum Likelihood and Maximum Entropy. *Optical Society of America*, vol. 62, (1972), 511–518.

[54] B. R. Frieden. Restoring with Maximum Likelihood and Maximum Entropy. *Journal of the Optical Society of America*, vol. 62, (1972), 511–518.

[55] J. E. Shore and R. W. Johnson. Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Cross-Entropy. *IEEE Transactions on Information Theory*, Vol. IT-26, No. 1, (1980), 26–37.

[56] J. E. Shore and R. W. Johnson. Properties of Cross-Entropy Minimization. *IEEE Transactions on Information Theory*, Vol. IT-27, No. 4, (1981), 472–482.

[57] J. E. Shore and R. W. Johnson. Comments and Corrections to "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy". *IEEE Transactions on Information Theory*, Vol. IT-26, No. 6, (1980), 784–785.

[58] S. F. Gull and J. Skilling, "The Maximum Entropy Method", *Indirect imaging, Proc., of International Symp.*, Sept 1983, Sydney, Cambridge University Press (1984) 267-278.

[59] R. Narayan and R. Nityananda, "Maximum Entropy - Flexibility Versus Fundamentalism", Indirect imaging, *Proc. of Int. Symp.*, Sept 1983, Sydney, Cambridge University Press (1984) 282-290.

[60] A. M. Djafari and G. Demoment, "Image Restoration and Reconstruction using Entropy as a Regularization Functional", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: Gary J. Erickson, C. R. Smith, Vol.2: Applications, Kluwer academic publishers, Dordrecht, Boston, London (1988) 341-355.

[61] E. T. Jaynes, "Prior Information and Ambiguity in Inverse Problems", *SIAM-AMS Proceedings*, Vol. 14 (1984) 151-166.

[62] S. Gull, "Bayesian Inductive Inference and Maximum Entropy", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: G. J. Erickson and C. Ray Smith, Vol. 1: foundations, Kluwer Academic Publishers, Dordrecht, the Netherlands (1988) 53-74.

[63] E. T. Jaynes, "The Relation of Bayesian and Maximum Entropy Methods", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: Gary J. Erickson, C. R. Smith, Kluwer academic publishers, Dordrecht, Boston, London(1988) 25-29.

[64] B. R. Frieden, " Fourier Methods in Probability", *Probaility, Statistical Optics, and Data Testing*, Springer-Verlag, Heidelberg, (1983) 70-96.

[50] S. J. Wernecke, "Two-Dimensional Maximum Entropy Reconstruction of Radio Brightness", *Radio Science*, Vol. 12, No. 5, (1977) 831 - 844.

[51] J. Skilling and S.F. Gull, " The Entropy of an Image", *SIAM-AMS proceedings, vol 14*, (1984) 167-189.

[52] S. F. Gull and J. Skilling, " The Entropy of an Image", *Maximum-Entropy and Bayesian Methods in Inverse Problems*, editors: C. R. Smith and W. T. Grandy. Jr., D. Reidel Publishing Company, Dordrecht, Holland (1985) 287-301.

[53] B. R. Frieden, " Estimating a Probability Law" in *Probaility, Statistical Optics, and Data Testing*, Springer-Verlag, Heidelberg, (1983) 264-286.

[54] B. R. Frieden, "Estimating Occurrence Laws with Maximum Probability, and the Transition to Entropic Estimators", *Maximum-Entropy and Bayesian Methods in Inverse Problems*, editors: C. R. Smith and W. T. Grandy. Jr., D. Reidel Publishing Company, Dordrecht, Holland (1985) 133-169.

[55] J. E. Shore and R. W. Johnson, "Axiomatic Derivation of the Priniciple of Maximum Entropy and the Principle of Cross-Entropy", *IEEE trans. Inform. Theory*, Vol. IT-26, No. 1, (1980) 26-37.

[56] J. E. Shore and R. W. Johnson, "Properties of Cross-Entropy Minimization", *IEEE trans. Inform. Theory*, Vol. IT-27, No. 4, (1980) 472-482.

[57] J. E. Shore and R. W. Johnson, "Inversion as Logical Inference-Theory and Applications of Maximum Entropy and Minimum Cross-Entropy", *SIAM-AMS Proceedings*, Vol. 14 (1984) 139-149.

[58] S. F. Gull and J. Skilling, "The Maximum Entropy Method", *Indirect imaging. Proc., of International Symp.*, Sept 1983, Sydney, Cambridge University Press (1984) 267-278.

[59] R. Narayan and R. Nityananda, "Maximum Entropy - Flexibility Versus Fundamentalism", Indirect imaging, *Proc. of Int. Symp.*, Sept 1983, Sydney, Cambridge University Press (1984) 282-290.

[60] A. M. Djafari and G. Demoment, "Image Restoration and Reconstruction using Entropy as a Regularization Functional", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: Gary J. Erickson, C. R. Smith, Vol.2: Applications, Kluwer academic publishers, Dordrecht, Boston, London (1988) 341-355.

[61] E. T. Jaynes, "Prior Information and Ambiguity in Inverse Problems", *SIAM-AMS Proceedings*, Vol. 14 (1984) 151-166.

[62] S. Gull, "Bayesian Inductive Inference and Maximum Entropy", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: G. J. Erickson and C. Ray Smith, Vol. 1: foundations, Kluwer Academic Publishers, Dordrecht, the Netherlands (1988) 53-74.

[63] E. T. Jaynes, "The Relation of Bayesian and Maximum Entropy Methods", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, editors: Gary J. Erickson, C. R. Smith, Kluwer academic publishers, Dordrecht, Boston, London(1988) 25-29.

[64] B. R. Frieden, " Fourier Methods in Probability", *Probaility, Statistical Optics, and Data Testing*, Springer-Verlag, Heidelberg, (1983) 70-96.

[65] E. Parzen, Modern Probability Theory and Its Applications, Wiley, Newyork (1968).

[66] P. E. Gill, W. Murray and M. H. Wright, Practical Optimization, Academic Press Inc., London, (1981).

[67] Nai-Long Wu, "Explicit and Implicit Solutions in the Maximum Entropy Method", *Indirect imaging, Proc. of International Symp.*, Sept 1983, Sydney, Cambridge University Press (1984) 303 - 308.

[68] S. W. Wilkins, "Statistical Geometry. I. A self-Consistent Approach to the Crystallographic Inversion Problem Based on Information Theory", *Acta cryst.*, A39, (1983) 47-60.

[69] S. W. Wilkins, "Statistical Geometry. II. Numerical Solution *via* the Single Pixel Equation", *Acta cryst.* A39, (1983) 892-896.

[70] X. Zhuang, E. Ostevold and R. M. Haralick, "A differential Equation Approach to Maximum Entropy Image Reconstruction", *IEEE trans. Acoust., Speech., Signal Processing*, Vol. ASSP-35, *No. 2*, (1987) 208-218.

[71] X. Zhuang, E. Ostevold and R. M. Haralick, "The Principle of Maximum Entropy in Image Recovery", *Image Recovery, Theory and Applications*, editor: H. Stark, Academic Press, (1987) 157-193.

[72] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, "Optimization by Simulated Annealing", *Science*, 220, (1983), 671-680.

[73] P. Carnevali, L. Coletti, S. Paternello, "Image Processing by Simulated Annealing", *Readings in Computer Vision: Issues, Problems, principles and Paradigms*, editors:

M. A. Fischler and O. Firschein, Morgan Kaufmann Publishers, Inc., Los Altos, CA 94022 (1987) 551-561.

[74] P. A. Stark, "Roots of Equations", *Introduction to Numerical Methods*, Macmillan Publishing Co., Inc. (1970) 68-122.

[75] Y. Jaluria, "Computer Methods for Engineering", Allyn and Bacon series in Engg, (1988).

[76] S. Y. Kung, " VLSI Array Processors" Prentice Hall , Englewood Cliffs, New Jersey (1988).

[77] R. Kasturi and J. Walkup, " Nonlinear Image Restoration in Signal-dependent Noise" , *Advances in Computer Vision and Image processing, A research annual, vol. 2: Image enhancement and Restoration*, editor: T. S. Huang, Jai press inc. connecticut, (1986) 167-212.

[78] T. G. Stockham, "Image Processing in the Context of a Visual Model", *Proceedings of the IEEE*, vol.60. No. 8, (1972) 828-842.

[79] J. S. Alkofer, "Tone Value Sample Selection in Digital Image Processing Method Employing Histogram Normalization", Patent, 4, 654,722, Mar. 31, 1987, U.S.A.

[80] F. Ratliff, "Mach bands: Quantitative Studies on Neural networks in the Retina", Holden-Day, Inc., San Francisco (1965).

[81] T. Cornsweet, "Visual Perception" Academic, New York (1970).

[82] L. S. Davis, "A Survey of Edge Detection Techniques", *Computer Graphics and Image Processing*, 4, (1975) 248-270.

[83] R.Haralick, "Digital step edges from zero crossing of 2nd directional derivatives" *Digital Image Processing and Analysis, Vol. 2: Digital Image Analysis*, editors: Rama Chellappa and A. A. Sawchuk, IEEE Computer Society Press, washington. D.C., (1985) 39-49.

[84] B. E. Bayer and P. G. Powell, "A Method for the Digital Enhancement of Unsharp, Grainy Photographic Images", *Adavances in Computer Vision and Image processing, A research annual, Vol. 2: Image enhancement and Restoration*, editor: T. S. Huang, Jai press inc. connecticut (1986) 31-88.

[85] J. A. B. Fortes, B. W. Wah, "Systolic Arrays: from Concept to Implementation (Guest editors introd. to spec. issue on Systolic Arrays)", *IEEE Computer*, (1987)12-17.

[86] W. B. Joerg, "Designing Dedicated Parallel Architectures with Petri Nets", *Proc. IEEE Pacific Rim Conf. on Communic., Computers and Signal Processing*, (1989),147-151.

[87] M. J. Flynn, "Some Computer Organizations and Their Effectiveness", *IEEE Trans. on Comp.* C-21, #9, (1972), 948-960.

[88] K. Hwang and F. A. Briggs, "Computer Architecture and Parallel Processing", McGraw Hill Inc., U. S. A.,(1984).

[89] M. J. Quinn, "Designing Efficient Algorithms for Parallel Computers", McGraw Hill Inc., U. S. A.,(1987).

[90] W. D. Hillis, G. L. Steele, "Data parallel algorithms", *CACM*, (1986), 1170-1183.

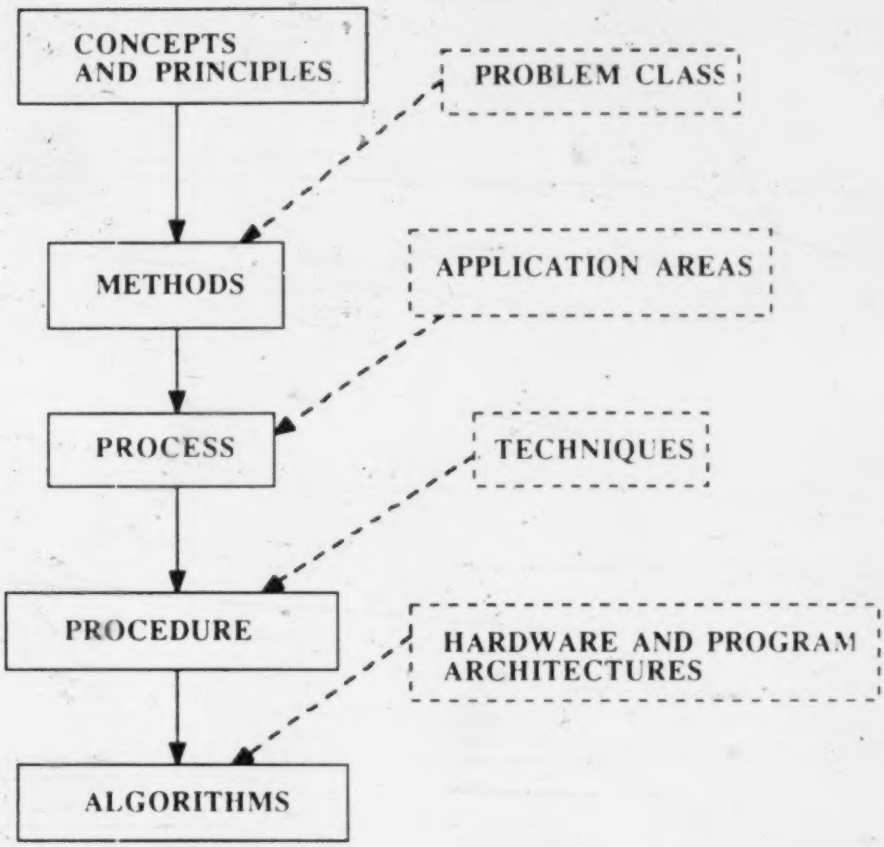
[91] System User's Manual, Myrias, PAMS. 2.3.0, (1989).

[92] B. W. Wah and C. V. Ramamoorthy, "Theory of Algorithms and Computation Complexity with Applications to Software Design", *Handbook of Software Engineering*, editors: C. R. Vick and C. V. Ramamoorthy, Van Nostrand Reinhold Company., New york (1984) 64-89.

[93] X. Qu and X. Li, "Parallel Template Matching Algorithms", *Proc. Int.conf on parallel processing*, (1988), 223-225.

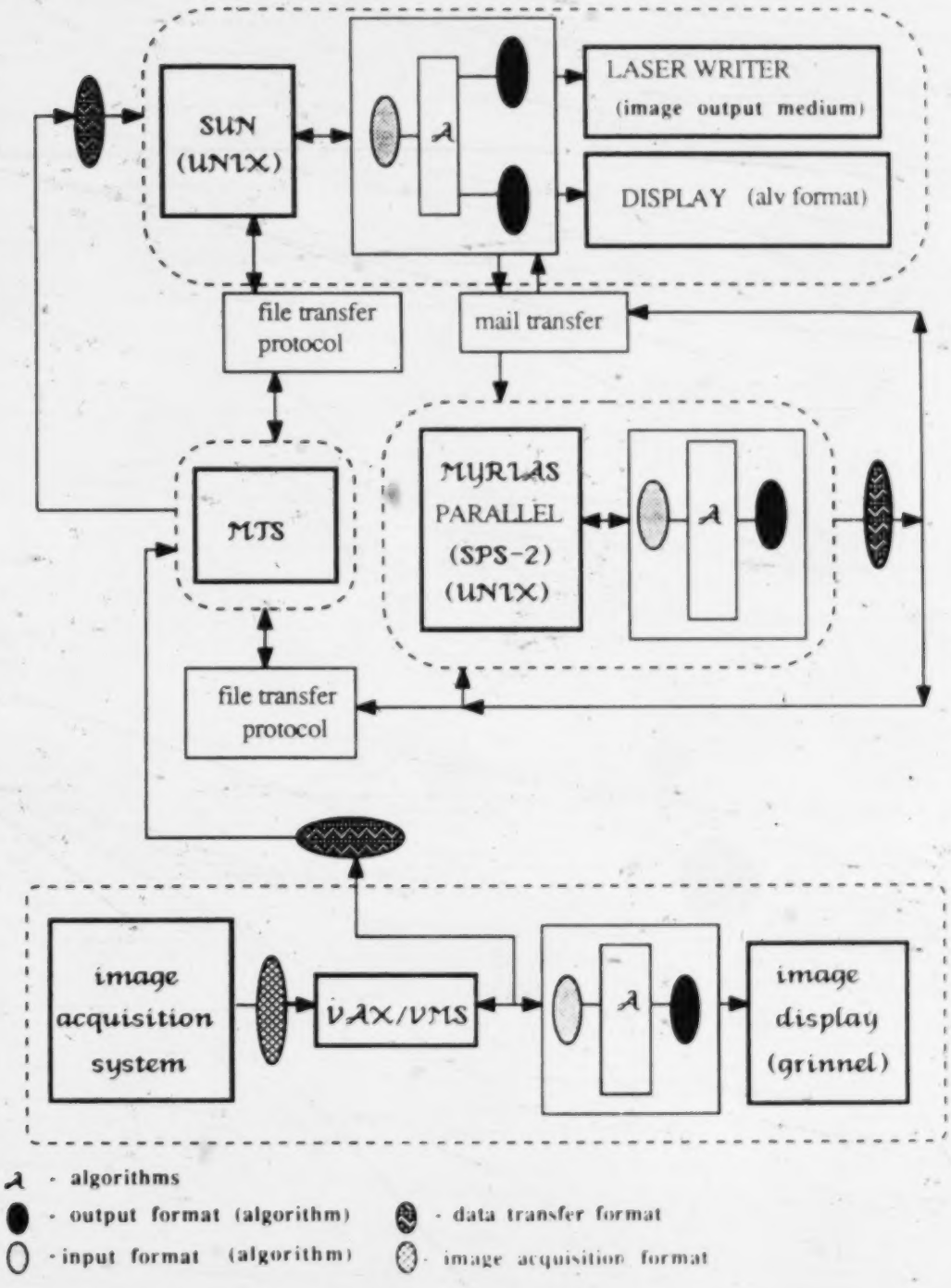
APPENDIX - A

Usage of different terms in the thesis



APPENDIX - B

Computing systems, input-output resources and migration paths



APPENDIX-C

The solution existence conditions for the single constraint *ME* method:

This section outlines the properties of J that determine the existence of a unique maximum point for the objective function in (3.4). For detailed discussions, readers are referred to [50], [70]. The purpose of this section is to study the consequence of removing the total constant intensity assumption on the solution conditions. Rewriting (3.4)

$$J(\mathbf{f}; \lambda_x) = S(\mathbf{f}) - \lambda_x \left(\frac{2Q(\mathbf{d},\mathbf{f})}{\sigma_x^2} - 1 \right) \tag{C-1}$$

The various terms follow the discussions of section 3.1.1. $S(\mathbf{f})$ is given by

$$S(\mathbf{f}) = -\sum_{j=1}^{n} \frac{f_j}{\sum_{k=1}^{n} f_k} \log \frac{f_j}{\sum_{k=1}^{n} f_k} \tag{C-2}$$

For convenience, $Q(\mathbf{d},\mathbf{f})$ is considered in its vector form as $Q(\mathbf{d},\mathbf{f}) = 1/2\ |A\mathbf{f} - \mathbf{d}|^2$ where

$$\mathbf{f} = \{f_1, \ldots\ldots, f_n\} \text{ and } A = [A_{jk}]_{m,n} \tag{C-3}$$

The following are the specific properties of J that determine the existence of its unique maximum point as a solution .

1. It is defined over a continuous domain:

$$\mathbf{D_f} \supset \{(f_1, \ldots\ldots, f_n);\ f_i > 0,\ i = 1, \ldots\ldots, n\} \tag{C-4}$$

2. It is twice differentiable with

$$\frac{\partial^2 J}{\partial f_j^2} = \frac{1 - \dfrac{f_j}{\sum_{k=1}^{n} f_k}}{f_j \left(\sum_{k=1}^{n} f_k \right)} \left(2 \frac{f_j}{\sum_{k=1}^{n} f_k} \log f_j - 1 \right) - \frac{\lambda_x}{\sigma_x^2} A^t A \tag{C-5}$$

and A^t is the transpose of the matrix A.

3. Its second derivative is negative definite everywhere when $\lambda_x \geq 0$ and $2(f_j / \Sigma f_k) \log f_j < 1$. The latter condition is due to the removal of total constant intensity assumption and can also be expressed as

$$\sum_{j=1}^{n} \frac{f_j}{\sum_{k=1}^{n} f_k} \log f_j < \frac{n}{2} . \qquad \text{(C-6)}$$

Further $J \rightarrow -\infty$, as $\| \mathbf{f} \| \rightarrow \infty$ (the symbol $\|.\|$ denotes the norm in $\mathbf{f}$ space). Based on the specific conditions listed above, and the detailed discussions in the cited references, J is a convex function with a unique and finite maximum point.

APPENDIX-D

Solution equation in Bayes' theorem approach

The derivation below is an alternative approach for deriving the solution equation starting from *flogf* form of entropy. The idea is to convey the mathematical steps involved in obtaining the solution equation. The discussions do not refer to the conceptual basis or the justifications for the usage of different quantities. Let us start with the *flogf* form of entropy and write the objective function from (3.4) as

$$J_1(\mathbf{f}) = S_1(\mathbf{f}) - \lambda_x \left(\frac{2Q(\mathbf{d},\mathbf{f})}{\sigma_x^2} - 1 \right) \tag{D-1}$$

where

$$S_1(\mathbf{f}) = \sum_{j=1}^{n} f_j \log f_j \tag{D-1-1}$$

The solution obtained by maximizing (D-1) is given by

$$f_j^1 = \exp\left\{ -1 - \frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d},\mathbf{f})}{\partial f_j} \right\}. \tag{D-2}$$

Dividing (D-2) by the sum total intensity of f_j^1

$$p_j^1 = \frac{1}{\sum_{i=1}^{n} f_i^1} \exp\left\{ -1 - \frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d},\mathbf{f})}{\partial f_j} \right\}. \tag{D-3}$$

Substituting (D-3) as a likelihood term in Bayes' theorem (equation (2.10)) and using the standard *plogp* form of entropy as prior probability distribution, the posterior probability becomes

$$p_j = \frac{1}{\sum_{i=1}^{n} f_i^1} \left[\exp(-S(\mathbf{f})) \exp\left\{ -1 - \frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d},\mathbf{f})}{\partial f_j} \right\} \right]. \tag{D-4}$$

Using (2.21), (D-4) can be written as

$$f_j = \frac{\sum_{k=1}^{n} f_k}{\sum_{i=1}^{n} f_i^d} \left[\exp(-S(\mathbf{f})) \exp\left\{ -1 - \frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d},\mathbf{f})}{\partial f_j} \right\} \right]. \qquad \text{(D-5)}$$

With suitable modifications, (D-5) can be written as

$$f_j = \exp\left(\log \sum_{k=1}^{n} f_k - S(\mathbf{f}) \right) \exp\left\{ -\frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d},\mathbf{f})}{\partial f_j} \right\} \left(\frac{\exp(-1)}{\sum_{i=1}^{n} f_i^d} \right) \qquad \text{(D-6)}$$

Using (3.2.2) and (3.2.3), (D-6) can be expressed as

$$f_j = q(\mathbf{f}) \exp\left\{ -\frac{2\lambda_x}{\sigma_x^2} \frac{\partial Q(\mathbf{d},\mathbf{f})}{\partial f_j} \right\} \left(\frac{\exp(-1)}{\sum_{i=1}^{n} f_i^d} \right). \qquad \text{(D-7)}$$

The first two terms in (D-7) together form the solution equation in (3.8) obtained by maximizing the *plogp* form of entropy. The extra mutiplication term is a constant that appears as a 'replacement term' for the eliminated $\sum_k f_k$ in the numerator of (3.8).

APPENDIX-E

True form of solution equation from its analytical form using β in (3.62)

Rewriting the analytical form of the solution equation given by (3.35)

$$\bar{f}_j^{(i)} = \left(1-\beta^{(i)}\right)\bar{f}_j^{(i-1)} + \beta^{(i)}\bar{f}_j^{(i)} . \qquad \text{(E-1)}$$

With $\beta^{(i)}$ given by (3.62), and $f_j^{(i)}$ given by (3.35.2), (E-1) becomes

$$\bar{f}_j^{(i)} = \left(1- \frac{1}{\lambda^{(i)}\sum_{k=1}^{n}\bar{f}_k^{(i)}}\right)\bar{f}_j^{(i-1)} + \frac{1}{\lambda^{(i)}\sum_{k=1}^{n}\bar{f}_k^{(i)}}\,\bar{q}^{(i)}\exp\left\{-\bar{\lambda}^{(i)}\frac{\partial\bar{Q}^{(i)}}{\partial f_j}\right\} . \qquad \text{(E-2)}$$

Substituting for $\bar{q}^{(i)}$ from (3.35.3) and rearranging the terms suitably

$$\bar{f}_j^{(i)} = \left(1- \frac{1}{\lambda^{(i)}\sum_{k=1}^{n}\bar{f}_k^{(i)}}\right)\bar{f}_j^{(i-1)} + \frac{1}{\lambda^{(i)}}\exp\left\{\left(\sum_{j=1}^{n}\bar{p}_j^{(i)}\log f_j^{(i)}\right) - \log\sum_{k=1}^{n}\bar{f}_k^{(i)}\right\}\exp\left\{-\bar{\lambda}^{(i)}\frac{\partial\bar{Q}^{(i)}}{\partial f_j}\right\} . \qquad \text{(E-3)}$$

Making use of the relationship in (3.2.1), (E-3) can also be written as

$$\bar{f}_j^{(i)}\sum_{k=1}^{n}\bar{f}_k^{(i)} = \bar{f}_j^{(i-1)}\sum_{k=1}^{n}\bar{f}_k^{(i)} + \frac{1}{\lambda^{(i)}}\left[\left\{\sum_{k=1}^{n}\bar{f}_k^{(i)}\exp\left\{-S\left(\bar{f}^{(i)}\right) - \bar{\lambda}^{(i)}\frac{\partial\bar{Q}^{(i)}}{\partial f_j}\right\}\right\} - \bar{f}_j^{(i-1)}\right] \qquad \text{(E-4)}$$

Substituting for $S(\bar{f})$ from (3.2.2) and comparing the result with the true form in (3.7), (E-4) becomes

$$\bar{f}_j^{(i)} = \bar{f}_j^{(i-1)} + \frac{1}{\lambda^{(i)}\sum_{k=1}^{n}\bar{f}_k^{(i)}}\left\{\bar{f}_{ij}^{(i)} - \bar{f}_j^{(i-1)}\right\} . \qquad \text{(E-5)}$$

Alternatively

$$\bar{f}_j^{(i)} = \left(1-\beta^{(i)}\right)\bar{f}_j^{(i-1)} + \beta^{(i)}\bar{f}_{ij}^{(i)} . \qquad \text{(E-6)}$$

where $\beta^{(i)}$ is given by

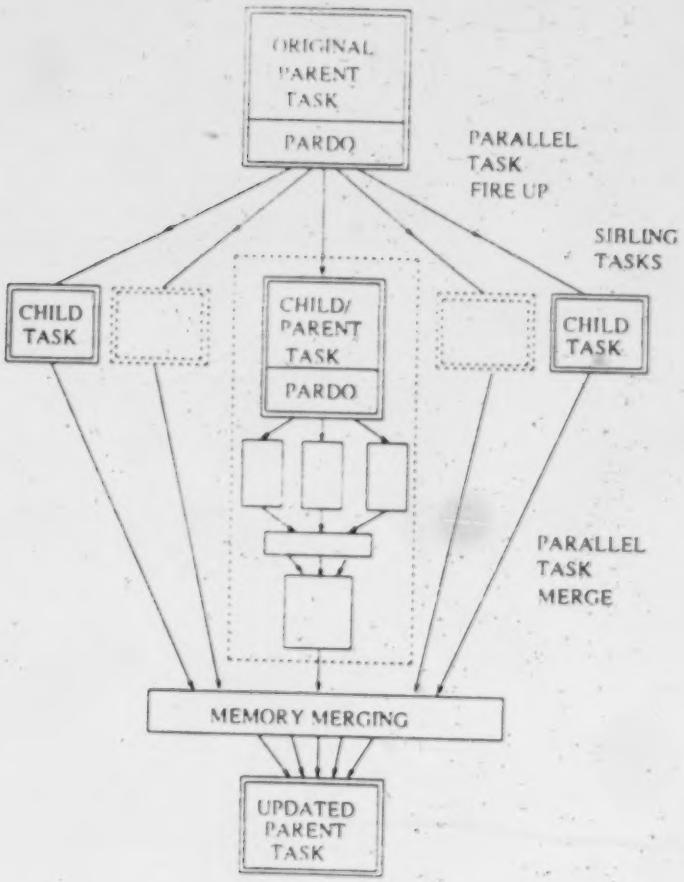
$$\beta^{(i)} = \frac{1}{\lambda^{(i)} \sum_{k=1}^{n} \bar{f}_k^{(i)}} . \qquad \text{(E-6-1)}$$

and

$$\bar{f}_{ij}^{(i)} = \bar{q}^{(i)} \exp\left\{ -\left(\sum_{k=1}^{n} \bar{f}_k^{(i)} \right) \bar{\lambda}^{(i)} \frac{\partial \bar{Q}^{(i)}}{\partial f_j} \right\} . \qquad \text{(E-6-2)}$$

APPENDIX - F *

Sequence of activities in the Myrias *pardo*



* includedwith permission from [91].

APPENDIX-G*

Run-time statistics in Myrias system

Statistics may be gathered during program execution. They are periodically sampled during program execution, and a summary is displayed at the end of program execution.

For information on how to interpret these statistics, see *User's Guide to Pardo*.

All basic statistics are domain-wide: the sum over all processors of individual processor statistics. For example, the *user cpu time* statistic is the sum over all processors of the user cpu time of each processor. Since cpu time is measured in seconds, the dimension of the summed statistic is *processor-seconds*. Typically, this sum exceeds the elapsed run-time of the program.

The displayed statistics are output to standard error unless the -f flag is used to redirect them to a file or program. Each statistic is identified by a two-letter code.

Statistics are subdivided into four categories: cpu time, paging, I/O, and program progress.

Cpu time statistics:

user cpu time (ut)	Time spent executing user code, summed over all processors in the domain. Measured in *processor-seconds*.
system cpu time (st)	Time spent executing system code, summed over all processors in the domain. Measured in *processor-seconds*.
idle cpu time (it)	Idle time for each cpu, summed over all processors in the domain. Measured in *processor-seconds*.
wait cpu time (wt)	Wait time for each cpu, summed over all processors in the domain. Measured in *processor-seconds*. Wait time is the time that a processor had a task resident that could not execute because it was waiting for a resource, typically a page.
primal time (pt)	Total time spent executing user code serially, in the primal task. Measured in *seconds*.
critical path time (ct)	Longest parallel execution path running user code. Measured in *seconds*.

Paging statistics:

pages in (pi)	Number of pages moved between processors to satisfy task page requests.
pages out (po)	Number of pages moved between processors as a result of over-commitment of memory on the processors initiating the move.
pages created (pc)	Number of pages created for each task to support the Myrias memory model.

I/O statistics:

svcs (sv)	Number of SVC server call instructions executed by the program.
svc bytes out (so)	Number of bytes transmitted by the program via SVC instructions.

* included with permission from [91].

APPENDIX-H

The recorded elapsed and idle times for the image convolution algorithm

number of processors	elapsed time (secs)	idle time (processor-secs)
1	80.1	0
2	41.3	0.3
4	21.0	0.5
8	10.8	1.5
10	8.7	2.4
16	5.5	1.8
24	3.8	5.6
32	2.8	5.1
50	2.2	21.1
60	1.9	23.2
64	1.7	17.3